

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 22:08:58 ; Search time 269.446 Seconds
(without alignments)
7888.487 Million cell updates/sec

Title: US-09-926-375b-7_COPY_12653_13951

Perfect score: 1299
Sequence: 1 atgaagcattcttaacc.....taccgcttcgattgttaa 1299

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 1202784 seqs, 81813859 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/prodata/1/ina/5A.COMB.seq: *
2: /cgn2_6/prodata/1/ina/5B.COMB.seq: *
3: /cgn2_6/prodata/1/ina/6A.COMB.seq: *
4: /cgn2_6/prodata/1/ina/6B.COMB.seq: *
5: /cgn2_6/prodata/1/ina/PCTUS.COMB.seq: *
6: /cgn2_6/prodata/1/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1282.2	98.7	1323	3	US-09-259-214-1 Sequence 1, Appl
2	1282.2	98.7	1323	3	US-09-318-528-1 Sequence 1, Appl
3	1282.2	98.7	1323	3	US-09-281-931-1 Sequence 1, Appl
4	1282.2	98.7	1323	4	US-09-580-515-1 Sequence 1, Appl
5	1257.4	96.8	1489	4	US-09-540-149A-9 Sequence 1, Appl
6	1170.2	90.1	1272	2	US-08-910-798-1 Sequence 1, Appl
7	63	4.8	1266	4	US-09-489-039A-341 Sequence 341, App
8	40.8	3.1	2868	4	US-09-710-794-4 Sequence 4, Appl
9	36.4	2.8	480	4	US-09-252-991A-15917 Sequence 15917, A
10	36.4	2.8	762	4	US-09-252-991A-15888 Sequence 15888, A
11	36.4	2.8	885	4	US-09-252-991A-15774 Sequence 15774, A
12	36.4	2.7	1066	4	US-09-252-991A-15804 Sequence 15804, A
13	35.4	2.7	3076	4	US-09-710-794-1 Sequence 1, Appl
14	34.6	2.7	2232	4	US-09-902-540-2830 Sequence 2830, Ap
15	34.6	2.7	14570	4	US-09-902-540-1012 Sequence 1012, Ap
16	34.4	2.6	4403765	3	US-09-103-840A-2 Sequence 2, Appl
17	34.4	2.6	4411529	3	US-09-103-840A-1 Sequence 1, Appl
18	34.2	2.6	505	4	US-09-621-976-15639 Sequence 15639, A
19	34	2.6	998	2	US-07-885-0898-5 Sequence 5, Appl
20	34	2.6	1044	4	US-09-252-991A-8907 Sequence 8907, Ap
21	34	2.6	1068	4	US-09-710-794-3 Sequence 3, Appl
22	34	2.6	2313	4	US-09-252-991A-8696 Sequence 8696, Ap
23	34	2.6	2706	4	US-09-252-991A-9082 Sequence 9082, Ap
24	33.8	2.6	12257	4	US-09-949-016-16384 Sequence 16384, A
25	33.6	2.6	505	4	US-09-621-976-15639 Sequence 15639, A
26	33.4	2.6	399	4	US-09-621-976-8976 Sequence 8976, Ap
27	33	2.5	2511	2	US-08-680-326-116 Sequence 116, App

C	28	33	2.5	35100	2	US-08-770-379-17	Sequence 17, Appl
C	29	33	2.5	35100	3	US-08-757-669A-17	Sequence 17, Appl
C	30	33	2.5	35100	3	US-09-230-371A-17	Sequence 17, Appl
C	31	33	2.5	69701	4	US-09-949-016-14187	Sequence 14187, A
C	32	33	2.5	73308	4	US-09-489-016-16326	Sequence 16326, A
C	33	32.8	2.5	1362	4	US-09-489-039A-2259	Sequence 2259, Ap
C	34	32.8	2.5	1389	4	US-09-489-039A-2538	Sequence 2538, Ap
C	35	32.6	2.5	1584	4	US-09-489-039A-497	Sequence 497, App
C	36	32.4	2.5	84425	4	US-09-949-016-17402	Sequence 17402, A
C	37	32.2	2.5	2430	4	US-09-489-039A-3732	Sequence 3732, Ap
C	38	32.2	2.5	12023	4	US-09-634-238-405	Sequence 405, App
C	39	32.2	2.5	20373	4	US-09-949-016-12302	Sequence 12302, A
C	40	32.2	2.5	20374	4	US-09-949-016-17086	Sequence 17086, A
C	41	32	2.5	1533	4	US-09-489-039A-4607	Sequence 4607, Ap
C	42	32	2.5	11454	4	US-09-949-016-14196	Sequence 14196, A
C	43	32	2.5	11459	4	US-09-949-016-12337	Sequence 12337, A
C	44	31.8	2.4	630	4	US-09-252-991A-6570	Sequence 6570, Ap
C	45	31.8	2.4	1656	4	US-09-252-991A-6620	Sequence 6620, Ap

ALIGNMENTS

RESULT 1
US-09-259-214-1
Sequence 1, Application US/09259214A
Patent No. 6110719
GENERAL INFORMATION:
APPLICANT: Kretz, Keith
TITLE OF INVENTION: NOVEL PHRYASE
FILE REFERENCE: DIVER1370-1
CURRENT APPLICATION NUMBER: US/09/259, 214A
CURRENT FILING DATE: 1999-03-01
EARLIER APPLICATION NUMBER: 08/910, 798
EARLIER FILING DATE: 1997-08-13
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 1323
TYPE: DNA
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(1320)
NAME/KEY: misc_feature
LOCATION: (1)...(1323)
OTHER INFORMATION: n = A,T,C or G
US-09-259-214-1

Query Match 98.7% Score 1282.2; DB 3; Length 1323;
Best Local Similarity 99.3% Pred. No. 0;
Matches 1287; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY	1	ATGAAAGCATCTTAATCCATTTTATCTCTCTGATTCGGTTAACCCGCAATCGCA	60
DB	1	ATGAAAGCATCTTAATCCATTTTATCTCTCTGATTCGGTTAACCCGCAATCGCA	60
QY	61	TTGCGTCAAGTGAAGCGGAGCTGAAGCTGAAAGTGTGATTTGTCAGTCTGATGT	120
DB	61	TTGCGTCAAGTGAAGCGGAGCTGAAGCTGAAAGTGTGATTTGTCAGTCTGATGT	120
QY	121	GTGCGTCTTCAACCAAGGCGACGCACTGATGCAAGATTCACCCGAGCGCATGGCCA	180
DB	121	GTGCGTCTTCAACCAAGGCGACGCACTGATGCAAGATTCACCCGAGCGCATGGCCA	180
QY	181	ACCTGGCGGTAATAACTGGGTTGGCTGACACCGCGCGTGGTGAATCGCTATCTC	240
DB	181	ACCTGGCGGTAATAACTGGGTTGGCTGACACCGCGCGTGGTGAATCGCTATCTC	240
QY	241	GACATTACCAAGCCGCGTCTGTGACCGACGGATTTGCTGCGAAAAAGGCTGCCG	300
DB	241	GACATTACCAAGCCGCGTCTGTGACCGACGGATTTGCTGCGAAAAAGGCTGCCG	300

```
QY 301 CAGTCTGCTCAGGTGCGCATTTTGTCTGATGTCGACGAGCGTAAACCCGTTAAACAGGCCAA 360
Db 301 CAGTCTGCTCAGGTGCGCATTTTGTCTGATGTCGACGAGCGTAAACCCGTTAAACAGGCCAA 360
QY 361 GCGTTCGCGCGCGGGCTGGCACCCTGACCTGTGCAATTAACCGTACATACCAGCAGATACG 420
Db 361 GCGTTCGCGCGCGGGCTGGCACCCTGACCTGTGCAATTAACCGTACATACCAGCAGATACG 420
QY 421 TCCAGTCCCGCATCCGTTATTTATCTCTTAAAACTGGCGTTTCCCACTGATTAACGCG 480
Db 421 TCCAGTCCCGCATCCGTTATTTATCTCTTAAAACTGGCGTTTCCCACTGATTAACGCG 480
QY 481 AACGTGACTGACCGCATCTCTGACGAGGGCAGAGGGGTCAATTGCTGACTTTACCGGGCAT 540
Db 481 AACGTGACTGACCGCATCTCTGACGAGGGCAGAGGGGTCAATTGCTGACTTTACCGGGCAT 540
QY 541 CGGCAAAACGCGGTTTGGCGAACGTGAAACGGAGCTTAAATTTTCCGCAATCAAACTGTGTC 600
Db 541 CGGCAAAACGCGGTTTGGCGAACGTGAAACGGAGCTTAAATTTTCCGCAATCAAACTGTGTC 600
QY 601 CTTAAACGTGAGAAACAGGACGAAAGCTGTCAATTAACGACGACATTAACATCGGAATC 660
Db 601 CTTAAACGTGAGAAACAGGACGAAAGCTGTCAATTAACGACGACATTAACATCGGAATC 660
QY 661 AAGGTGAGCGCGGACAAATGTTCTATTAACCGGTGCGGTAAAGCTCCGCAATCAATGCTGAG 720
Db 661 AAGGTGAGCGCGGACAAATGTTCTATTAACCGGTGCGGTAAAGCTCCGCAATCAATGCTGAG 720
QY 721 GAGATATTTCTCTGCAACAGCAAGGAAATGCGGAAACCGGGTGGGGAAGATCAC 780
Db 721 GAGATATTTCTCTGCAACAGCAAGGAAATGCGGAAATGCGGAAACCGGGTGGGGAAGATCAC 780
QY 781 GATTCAACACAGTGAACAACCTTGTGAAGTTGCAATACGCGCAATTTTATTTGCTACAA 840
Db 781 GATTCAACACAGTGAACAACCTTGTGAAGTTGCAATACGCGCAATTTTATTTGCTACAA 840
QY 841 CGGACGCCGAGAGTTTCCCGCAGCCGCGCAACCCCGTTATTAAGTTTGAATCAAGACAGG 900
Db 841 CGGACGCCGAGAGTTTCCCGCAGCCGCGCAACCCCGTTATTAAGTTTGAATCAAGACAGG 900
QY 901 TTGACCGCCCATCAACGCAAAACAGGGGTATGCTGATCAATTAACCTTCAAGTCTG 960
Db 901 TTGACCGCCCATCAACGCAAAACAGGGGTATGCTGATCAATTAACCTTCAAGTCTG 960
QY 961 TTTATGCGCGGACACGATTAATCTGCGCAATCTGCGCGGCACTGAGCTCAACTG 1020
Db 961 TTTATGCGCGGACACGATTAATCTGCGCAATCTGCGCGGCACTGAGCTCAACTG 1020
QY 1021 ACGCTTCCCGGTACGCGGATTAACAGCGCGCAGGTGTAATCTGCGGTTTGAACGCTG 1080
Db 1021 ACGCTTCCCGGTACGCGGATTAACAGCGCGCAGGTGTAATCTGCGGTTTGAACGCTG 1080
QY 1081 CGTCCGCTAAGCGATTAACAGCGAGTGAATCAAGTTTCGTGCTTCCAGACTTTACG 1140
Db 1081 CGTCCGCTAAGCGATTAACAGCGAGTGAATCAAGTTTCGTGCTTCCAGACTTTACG 1140
QY 1141 CAGATGCGATTAATAAGCGCGCTGTCTTAATTAATACCGCGCGGAGAGGTGAATCGACC 1200
Db 1141 CAGATGCGATTAATAAGCGCGCTGTCTTAATTAATACCGCGCGGAGAGGTGAATCGACC 1200
QY 1201 CTGGCAGAGTGAATAAGGAGAAATGGCAGGGCATGTTGTTGTTGGCAGGTTTACGCA 1260
Db 1201 CTGGCAGAGTGAATAAGGAGAAATGGCAGGGCATGTTGTTGTTGGCAGGTTTACGCA 1260
QY 1261 ATCGTGAATGAACACGATACCGGCTTGGCAGTTTG 1296
Db 1261 ATCGTGAATGAACACGATACCGGCTTGGCAGTTTG 1296
```

RESULT 2
US-09-318-528-1
; Sequence 1, Application US/09318528
; Patent No. 6183740

```
GENERAL INFORMATION:  
; APPLICANT: Kretz, Keith  
; TITLE OF INVENTION: NOVEL PHYLASE  
; FILE REFERENCE: 09/010,029,003  
; CURRENT APPLICATION NUMBER: US/09/318,528  
; CURRENT FILING DATE: 1999-05-25  
; EARLIER APPLICATION NUMBER: 09/291,931  
; EARLIER FILING DATE: 1999-04-13  
; EARLIER APPLICATION NUMBER: 08/910,798  
; EARLIER FILING DATE: 1997-08-13  
; EARLIER APPLICATION NUMBER: 09/259,214  
; EARLIER FILING DATE: 1999-03-01  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO: 1  
; LENGTH: 1323  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(1320)  
; NAME/KEY: misc_feature  
; LOCATION: (1)...(1323)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-318-528-1  
Query Match 98.7%; Score 1282.2; DB 3; Length 1323;  
Best Local Similarity 99.3%; Pred.No. 0;  
Matches 1287; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
QY 1 ATGAAAGCATCTTAATCCATTTTATCTCTTGAATTCGTTAAACCCGCAATCTGCA 60  
Db 1 ATGAAAGCATCTTAATCCATTTTATCTCTTGAATTCGTTAAACCCGCAATCTGCA 60  
QY 61 TTGCTCAGAGTGAACCGGACGCTGAAGCTGGAAGTGTGTGATTTGCTATGCT 120  
Db 61 TTGCTCAGAGTGAACCGGACGCTGAAGCTGGAAGTGTGTGATTTGCTATGCT 120  
QY 121 GTGCGTCTCCCAACCAAGGCCACGCACTGATGACGAGATGTCACCCGACGATGGCA 180  
Db 121 GTGCGTCTCCCAACCAAGGCCACGCACTGATGACGAGATGTCACCCGACGATGGCA 180  
QY 181 ACCTGGCGGTAAACTGGGTTGCTGACACCGCGCGGTGTGAGCTTAATCGCTATCTC 240  
Db 181 ACCTGGCGGTAAACTGGGTTGCTGACACCGCGCGGTGTGAGCTTAATCGCTATCTC 240  
QY 241 GGAATTAACCAACGCCAGGCTTGTGATGCTGACGATTCCTGCGCAAAAGGGCTGCGG 300  
Db 241 GGAATTAACCAACGCCAGGCTTGTGATGCTGACGATTCCTGCGCAAAAGGGCTGCGG 300  
QY 301 CAGTCTGCTCAGGTGCGCATTTTGTCTGATGTCGACGAGCGTAAACAGGCGAA 360  
Db 301 CAGTCTGCTCAGGTGCGCATTTTGTCTGATGTCGACGAGCGTAAACAGGCGAA 360  
QY 361 GCGTTCGCGCGCGGGCTGGCACCCTGACCTGTGCAATTAACCGTACATACCAGCAGATACG 420  
Db 361 GCGTTCGCGCGCGGGCTGGCACCCTGACCTGTGCAATTAACCGTACATACCAGCAGATACG 420  
QY 421 TCCAGTCCCGATCCGTTATTTATCTCTTAAAACTGGCGTTTCCCACTGATTAACGCG 480  
Db 421 TCCAGTCCCGATCCGTTATTTATCTCTTAAAACTGGCGTTTCCCACTGATTAACGCG 480  
QY 481 AACGTGACTGACCGCATCTCTGACGAGGGCAGAGGGGTCAATTGCTGACTTTACCGGGCAT 540  
Db 481 AACGTGACTGACCGCATCTCTGACGAGGGCAGAGGGGTCAATTGCTGACTTTACCGGGCAT 540  
QY 541 CGGCAAAACGCGGTTTGGCGAACGTGAAACGGAGCTTAAATTTTCCGCAATCAAACTGTGTC 600  
Db 541 CGGCAAAACGCGGTTTGGCGAACGTGAAACGGAGCTTAAATTTTCCGCAATCAAACTGTGTC 600  
QY 601 CTTAAACGTGAGAAACAGGACGAAAGCTGTCAATTAACGACGACATTAACATCGGAATC 660  
Db 601 CTTAAACGTGAGAAACAGGACGAAAGCTGTCAATTAACGACGACATTAACATCGGAATC 660
```

Qy	661	AAAGTGAAGCCGGCAACAAGTCTCAATTAACCGGTGGGATGAAGCTTCGATCAATGCTGAC	720
Db	661	AAGGTGAAGCCGGCAACAAGTCTCAATTAACCGGTGGGATGAAGCTTCGATCAATGCTGAC	720
Qy	721	GAGATATTTCCTCTGCAACAAGCAACAGGAAATGCCGAGACCGGGGTGGGAAAGATCAAC	780
Db	721	GAGATATTTCCTCTGCAACAAGCAACAGGAAATGCCGAGACCGGGGTGGGAAAGATCAAC	780
Qy	781	GATTCAACACAGTGGAAACACTTGTGTAAGTTTGGCTAAACGGGCAATTTTATTGCTACAA	840
Db	781	GATTCAACACAGTGGAAACACTTGTGTAAGTTTGGCTAAACGGGCAATTTTATTGCTACAA	840
Qy	841	CGCACGCCAGAGGTTGGCCCGGACCGCGCACCCCGTTAATTAGATTGTCAAGAAGAGG	900
Db	841	CGCACGCCAGAGGTTGGCCCGGACCGCGCACCCCGTTAATTAGATTGTCAAGAAGAGG	900
Qy	901	TTGACGCCCCATCCACCGCAAAAACAGCGGATGAGTGTGACATTAACCACTTCAGTGGCTG	960
Db	901	TTGACGCCCCATCCACCGCAAAAACAGCGGATGAGTGTGACATTAACCACTTCAGTGGCTG	960
Qy	961	TTTATCGCCGACACGATTACTAATCTGGCAATCTTCGGCGCGCACTGGAAGTCAACTGG	1020
Db	961	TTTATCGCCGACACGATTACTAATCTGGCAATCTTCGGCGCGCACTGGAAGTCAACTGG	1020
Qy	1021	AACGCTCCCGGTCACCGGATTAACAAGCGCGGACAGGATGGGAACTGGTGTGTAAGAGCTGG	1080
Db	1021	AACGCTCCCGGTCACCGGATTAACAAGCGCGGACAGGATGGGAACTGGTGTGTAAGAGCTGG	1080
Qy	1081	CGTCGGCTAAGCGATTAACAGCCAGTGGATTCAAGTTTCGCTGGTCTTCACAGCTTTACAG	1140
Db	1081	CGTCGGCTAAGCGATTAACAGCCAGTGGATTCAAGTTTCGCTGGTCTTCACAGCTTTACAG	1140
Qy	1141	CAGATGCGTGATTAACAGCCGCTGTTCATTAATTAAGCCGCCCGGAGAGGTGAACCTGAC	1200
Db	1141	CAGATGCGTGATTAACAGCCGCTGTTCATTAATTAAGCCGCCCGGAGAGGTGAACCTGAC	1200
Qy	1201	CTGGCAGAGTGTGAAGAGCGGAATGCGACGAGCATGTGTCTTGGCAGAGTTTACGCAA	1260
Db	1201	CTGGCAGAGTGTGAAGAGCGGAATGCGACGAGCATGTGTCTTGGCAGAGTTTACGCAA	1260
Qy	1261	ATCGTGAATGAAGACAGCATACCCGCTTGCAGTTTG	1296
Db	1261	ATCGTGAATGAAGACAGCATACCCGCTTGCAGTTTG	1296

```

RESULT 3
US-09-291-931-1
; Sequence 1, Application US/09291931A
; Patent No. 6190897
; GENERAL INFORMATION:
; APPLICANT: Kretz, Keith
; TITLE OF INVENTION: NOVEL PHYTASE
; FILE REFERENCE: 09010/029003
; CURRENT APPLICATION NUMBER: US/09/291,931A
; EARLIER FILING DATE: 1999-04-13
; EARLIER APPLICATION NUMBER: 08/910,798
; EARLIER FILING DATE: 1997-08-13
; EARLIER APPLICATION NUMBER: 09/259,214
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1323
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1320)
; NAME/KEY: misc_feature
; LOCATION: (1)...(1323)
; OTHER INFORMATION: n = A,T,C or G
US-09-291-931-1

```

Query Match	98.7%;	Score 1282.2;	DB 3;	Length 1323;
Best Local Similarity	99.3%;	Pred. No. 0;		
Matches 1287;	Conservative 0;	Mismatches 9;	Indels 0;	Gaps 0;
QY	1	ATGAAAGCCATCTTAAATCCCATTTTATCTCTTCTGTATTCCTCGTTAAACCCCGCAATCTGCA	60	
Db	1	ATGAAAGCCGATCTTAAATCCCATTTTATCTCTTCTGTATTCCTCGTTAAACCCCGCAATCTGCA	60	
QY	61	TTGGCTCAGAGTGAAGCCCGAGAGCTGAAGCTGGAAGATGTGTGATTTGTCACTCGCATGT	120	
Db	61	TTGGCTCAGAGTGAAGCCCGAGAGCTGAAGCTGGAAGATGTGTGATTTGTCACTCGCATGT	120	
QY	121	GTGGGTGCTCCAAACCAAGGCCACGCAACTGATGAGGATGTCAACCCGACGCAATGCGCA	180	
Db	121	GTGGGTGCTCCAAACCAAGGCCACGCAACTGATGAGGATGTCAACCCGACGCAATGCGCA	180	
QY	181	ACCTGGCCGGTAAACCTGGTGTGGCTGTACACCGCGCGGTGTGAGCTAAATCCGCTTATCTC	240	
Db	181	ACCTGGCCGGTAAACCTGGTGTGGCTGTACACCGCGMGSTGTGAGCTAAATCCGCTTATCTC	240	
QY	241	GGAATTAAACCAAGCCGACGCGTCTGTGTAACCGCATTTGCTGGGAAAAAGGCGCTGGCCG	300	
Db	241	GGAATTAAACCAAGCCGACGCGTCTGTGTAACCGCATTTGCTGGGAAAAAGGCGCTGGCCG	300	
QY	301	CAGTCTGTGTCAGGTGCGGATTTATTTGCTGATGTGACGACGCGTACCCGTTAAACAGGCGAA	360	
Db	301	CAGTCTGTGTCAGGTGCGGATTTATTTGCTGATGTGACGACGAGGTATCCGTTAAACAGGCGAA	360	
QY	361	GCCTTGCGCGCGCGGGCGTGGACCTCGTACCTGTGCAATTAACCGTACATACCAGGCGAGATACG	420	
Db	361	GCCTTGCGCGCGCGGGCGTGGACCTCGTACCTGTGCAATTAACCGTACATACCAGGCGAGATACG	420	
QY	421	TCCAGTCCCGATCCGTTATTTAACTCTTAATAAATGCGGCTTTGSCAACTGGATPACGCG	480	
Db	421	TCCAGTCCCGATCCGTTATTTAACTCTTAATAAATGCGGCTTTGSCAACTGGATPACGCG	480	
QY	481	AACGTGACTGACCGGATTCCTCAAGACGGGCGAGAGGTCATATTCGTGACTTTAACGGGCGAT	540	
Db	481	AACGTGACTGACCGGATTCCTCAAGACGGGCGAGAGGTCATATTCGTGACTTTAACGGGCGAT	540	
QY	541	CGGCAAAACGGGGTTTCGCGAACTGGAACGGGTGCTTAATTTTCGCAATCAAACTTGTGC	600	
Db	541	CGGCAAAACGGGGTTTCGCGAACTGGAACGGGTGCTTAATTTTCGCAATCAAACTTGTGC	600	
QY	601	CTTAAACGTGAGAAACAGAGAGAAAGCTGTTGATTTAACGACGCAATTAACATCGAAATCTC	660	
Db	601	CTTAAACGTGAGAAACAGAGAGAAAGCTGTTGATTTAACGACGCAATTAACATCGAAATCTC	660	
QY	661	AAGGTGAGCGCGGACATGTCTCATTTAACCGGTGCGGTAAAGCCGTGCAATCAATGCTAGC	720	
Db	661	AAGGTGAGCGCGGACATGTCTCATTTAACCGGTGCGGTAAAGCCGTGCAATCAATGCTAGC	720	
QY	721	GAGATATTTCCTCGCAACAAGCAACAGGGAATGCGGAGCGGGGTGGGGAAGATCAAC	780	
Db	721	GAGATATTTCCTCGCAACAAGCAACAGGGAATGCGGAGCGGGGTGGGGAAGATCAAC	780	
QY	781	GATTCAACCAAGTGAACACCTTGCTTAAGTTGCAATAACGCGCAATTTTATTTGCTACAA	840	
Db	781	GATTCAACCAAGTGAACACCTTGCTTAAGTTGCAATAACGCGCAATTTTATTTGCTACAA	840	
QY	841	CGCAGCGCAAGGTGTGCGCGGACGCGGCAACCCCGTTATTGATTTGATCATGCGACG	900	
Db	841	CGCAGCGCAAGGTGTGCGCGGACGCGGCAACCCCGTTATTGATTTGATCATGCGACG	900	
QY	901	TTGAACGCCCATTCACCGCAAAAACAGGCGTATGTGTGATTAACCATTCAGTACTG	960	
Db	901	TTGAACGCCCATTCACCGCAAAAACAGGCGTATGTGTGATTAACCATTCAGTACTG	960	
QY	961	TTTATTCGCGGCAACGATTAATCTGTGGCAAAATCTCGCGCGGCGCACTGGAGACTCAACTGG	1020	
Db	961	TTTATTCGCGGCAACGATTAATCTGTGGCAAAATCTCGCGCGGCGCACTGGAGACTCAACTGG	1020	

QY 1021 ACGTTTCCCGTACAGCCGGAATACAGCCGCCAGGTGGTGAACGTGTTTGAACGCTGG 1080
|
|
|
Db 1021 ACGTTTCCCGTACAGCCGGAATACAGCCGCCAGGTGGTGAACGTGTTTGAACGCTGG 1080
|
|
|
QY 1081 CGTCCGCTAAGCGATTAACAGCCAGTGAATTCAGTTTCGTGGTCTTCAGACTTTACG 1140
|
|
|
Db 1081 CGTCCGCTAAGCGATTAACAGCCAGTGAATTCAGTTTCGTGGTCTTCAGACTTTACG 1140
|
|
|
QY 1141 CAGATGCTGATTAACCGCCGCTGCTATTAATAGCCGCCCGGAGAGGTGAACCTGAC 1200
|
|
|
Db 1141 CAGATGCTGATTAACCGCCGCTGCTATTAATAGCCGCCCGGAGAGGTGAACCTGAC 1200
|
|
|
QY 1201 CTGGAGAGATGTAAGAGCGGAATGCGAGGACATGTGTGTTGTTGAGGTTTACGCA 1260
|
|
|
Db 1201 CTGGAGAGATGTAAGAGCGGAATGCGAGGACATGTGTGTTGTTGAGGTTTACGCA 1260
|
|
|
QY 1261 ATCGTAATGAAGACGACATACCCGCTTGCACTTG 1296
|
|
|
Db 1261 ATCGTAATGAAGACGACATACCCGCTTGCACTTG 1296
|
|
|

RESULT 4

US-09-580-515-1
; Sequence 1, Application US/09580515
; Patent No. 6720014
; GENERAL INFORMATION:
; APPLICANT: Kretz, Keith
; TITLE OF INVENTION: NOVEL PHYTASE
; FILE REFERENCE: 09010/029003
; CURRENT APPLICATION NUMBER: US/09/580,515
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: 09/318,528
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: 09/291,931
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: 08/910,798
; PRIOR FILING DATE: 1997-08-13
; PRIOR APPLICATION NUMBER: 09/259,214
; PRIOR FILING DATE: 1999-03-01
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1323
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1320)
; NAME/KEY: misc feature
; LOCATION: (1)...(1323)
; OTHER INFORMATION: n = A,T,C or G
US-09-580-515-1

Query Match 98.7%; Score 1282.2; DB 4; Length 1323;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 1287; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 ATGAAAGCATTTAATCCCATTTTATCTCTTGATTCGTTAACCCCGCAATCTGCA 60
|
|
|
Db 1 ATGAAAGCATTTAATCCCATTTTATCTCTTGATTCGTTAACCCCGCAATCTGCA 60
|
|
|
QY 61 TTTCGCTCAGAGTGAAGCCGAGCTGAAGCTGGAAGTGTGTGATTTGCTCATGTG 120
|
|
|
Db 61 TTTCGCTCAGAGTGAAGCCGAGCTGAAGCTGGAAGTGTGTGATTTGCTCATGTG 120
|
|
|
QY 121 GTGCGTGTCTCAACCAAGCCAGCACTGATGCAAGATGTCAACCCAGAGCGATGCGCA 180
|
|
|
Db 121 GTGCGTGTCTCAACCAAGCCAGCACTGATGCAAGATGTCAACCCAGAGCGATGCGCA 180
|
|
|
QY 181 ACCTGGCCGGTAAACTGGGTTGGCTGACACCGCGGTTGGTGAAGCTTAATCGCTATCTC 240
|
|
|
Db 181 ACCTGGCCGGTAAACTGGGTTGGCTGACACCGCGGTTGGTGAAGCTTAATCGCTATCTC 240
|
|
|

QY 241 GAACTTACCAACGCGACAGCTCTGTAGCCGACGGAATGTGCGGAAAAAGGCTGCCG 300
|
|
|
Db 241 GAACTTACCAACGCGACAGCTCTGTAGCCGACGGAATGTGCGGAAAAAGGCTGCCG 300
|
|
|
QY 301 CAGTGTGCTCAGTGGCGATTAATGCTGATGTCAACAGAGCTTACCCGTAACAGGCGAA 360
|
|
|
Db 301 CAGTGTGCTCAGTGGCGATTAATGCTGATGTCAACAGAGCTTACCCGTAACAGGCGAA 360
|
|
|
QY 361 GCGTTCCGCGCGCGGCTGGCACTGACGTGCAATTAACCGTACACAGGAGATAG 420
|
|
|
Db 361 GCGTTCCGCGCGCGGCTGGCACTGACGTGCAATTAACCGTACACAGGAGATAG 420
|
|
|
QY 421 TCAGTCCCGATTCGTTATTTAATCTCTTAATAAACTGCGTTTGCAACTGGAATACGCG 480
|
|
|
Db 421 TCAGTCCCGATTCGTTATTTAATCTCTTAATAAACTGCGTTTGCAACTGGAATACGCG 480
|
|
|
QY 481 AACGTACTGAACGCGATCTTCAACAGGCGAGAGGTCATTTGCTGACTTTACGGGCAAT 540
|
|
|
Db 481 AACGTACTGAACGCGATCTTCAACAGGCGAGAGGTCATTTGCTGACTTTACGGGCAAT 540
|
|
|
QY 541 CGGCAAAAGCGGTTTCGGAACCTGGAACGGGTCTTAATTTTCGCAATCAAACTTGTC 600
|
|
|
Db 541 CGGCAAAAGCGGTTTCGGAACCTGGAACGGGTCTTAATTTTCGCAATCAAACTTGTC 600
|
|
|
QY 601 CTTAAACGTGAAGAAACAGAGCAAGAACTGTTCAATTAACGAGCAATTAACATCGGAATC 660
|
|
|
Db 601 CTTAAACGTGAAGAAACAGAGCAAGAACTGTTCAATTAACGAGCAATTAACATCGGAATC 660
|
|
|
QY 661 AAGGTAGCGCGCAATGTCTATTAAACGGTGGGTTAGGCTTCGATCAATGCTGACG 720
|
|
|
Db 661 AAGGTAGCGCGCAATGTCTATTAAACGGTGGGTTAGGCTTCGATCAATGCTGACG 720
|
|
|
QY 721 GAGATATTTCTCTGCAACAGCAAGGAAATGCGGAGCGGGTGGGAAAGATCAAC 780
|
|
|
Db 721 GAGATATTTCTCTGCAACAGCAAGGAAATGCGGAGCGGGTGGGAAAGATCAAC 780
|
|
|
QY 781 GATTACACACAGTGAACACCTTGCTAATGCTGAATACGCGCAATTTATTTGCTACAA 840
|
|
|
Db 781 GATTACACACAGTGAACACCTTGCTAATGCTGAATACGCGCAATTTATTTGCTACAA 840
|
|
|
QY 841 CGCAGCGCAAGGTTGCCCGCAGCGCCGACCCCGTATTGATTTGATATGACAGCAGC 900
|
|
|
Db 841 CGCAGCGCAAGGTTGCCCGCAGCGCCGACCCCGTATTGATTTGATATGACAGCAGC 900
|
|
|
QY 901 TTGAGCGCCCATCCACCGCAAAAACAGGCGTATGATGATTAACCACTTCACTGCTG 960
|
|
|
Db 901 TTGAGCGCCCATCCACCGCAAAAACAGGCGTATGATGATTAACCACTTCACTGCTG 960
|
|
|
QY 961 TTTATCGCGGACAGATTAATCTGCAATCTCGCGGCGCACTGAGCTCAACTGG 1020
|
|
|
Db 961 TTTATCGCGGACAGATTAATCTGCAATCTCGCGGCGCACTGAGCTCAACTGG 1020
|
|
|
QY 1021 ACGTTTCCCGTACAGCCGGAATACAGCCGCCAGGTGTGAACGTGTGTTGAAGCTGG 1080
|
|
|
Db 1021 ACGTTTCCCGTACAGCCGGAATACAGCCGCCAGGTGTGAACGTGTGTTGAAGCTGG 1080
|
|
|
QY 1081 CGTCCGCTAAGCGATTAACAGCCAGTGAATTCAGTTTCGTGGTCTTCAGACTTTACG 1140
|
|
|
Db 1081 CGTCCGCTAAGCGATTAACAGCCAGTGAATTCAGTTTCGTGGTCTTCAGACTTTACG 1140
|
|
|
QY 1141 CAGATGCTGATTAACCGCCGCTGCTATTAATAGCCGCCCGGAGAGGTGAACCTGAC 1200
|
|
|
Db 1141 CAGATGCTGATTAACCGCCGCTGCTATTAATAGCCGCCCGGAGAGGTGAACCTGAC 1200
|
|
|
QY 1201 CTGGAGAGATGTAAGAGCGGAATGCGAGGACATGTGTGTTGTTGAGGTTTACGCA 1260
|
|
|
Db 1201 CTGGAGAGATGTAAGAGCGGAATGCGAGGACATGTGTGTTGTTGAGGTTTACGCA 1260
|
|
|
QY 1261 ATCGTAATGAAGACGACATACCCGCTTGCACTTG 1296
|
|
|
Db 1261 ATCGTAATGAAGACGACATACCCGCTTGCACTTG 1296
|
|
|

RESULT 5
US-09-540-149A-9
Sequence 9, Application US/09540149A
Patent No. 6511699
GENERAL INFORMATION:
APPLICANT: lei, Xingen
TITLE OF INVENTION: ENZYMES WITH IMPROVED PHYTASE ACTIVITY
FILE REFERENCE: 19603/2791
CURRENT APPLICATION NUMBER: US/09/540,149A
CURRENT FILING DATE: 2000-03-31
PRIOR APPLICATION NUMBER: 60/127,032
PRIOR FILING DATE: 1999-03-31
NUMBER OF SEQ ID NOS: 9
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 9
LENGTH: 1489
TYPE: DNA
ORGANISM: Escherichia coli
US-09-540-149A-9

Query Match 96.8%; Score 1257.4; DB 4; Length 1489;
Best Local Similarity 98.0%; Pred. No. 0;
Matches 1273; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 1 ATGAAAGCCATCTTAAATCCATTTTATCTCTTCTGATTCGGTTAAACCCCGCATCTGCA 60
DB 182 ATGAAAGCCATCTTAAATCCATTTTATCTCTTCTGATTCGGTTAAACCCCGCATCTGCA 241
QY 61 TTGGCTCAGATGAGCGCGAGCTGAAGCTGAAAGTGTGATGTCAGTCTGATGT 120
DB 242 TTGGCTCAGATGAGCGCGAGCTGAAGTGTGATGTCAGTCTGATGT 301
QY 121 GTGCTGCTTCAACCAAGGCCAAGCACTGATGAGGATGTCAACCCGAGCGATGCGCA 180
DB 302 GTGCTGCTTCAACCAAGGCCAAGCACTGATGAGGATGTCAACCCGAGCGATGCGCA 361
QY 181 ACCCTGCGGTTAAATCTGAGTGTGCTGACACCGCGCGTGTGATGCTAATCTCTATCTC 240
DB 362 ACCCTGCGGTTAAATCTGAGTGTGCTGACACCGCGCGTGTGATGCTAATCTCTATCTC 421
QY 241 GGAATTAACCAAGGCCAAGCGCTGTGAGCGGATGTCGGGGAAGGAGGCTGCCG 300
DB 422 GGAATTAACCAAGGCCAAGCGCTGTGAGCGGATGTCGGGGAAGGAGGCTGCCG 481
QY 301 CAGCTGTGATGAGTGTGATGATGCTGATGTCAGCAGGATGACCCGTAAACAGCGGAA 360
DB 482 CAGCTGTGATGAGTGTGATGATGCTGATGTCAGCAGGATGACCCGTAAACAGCGGAA 541
QY 361 GCTTGTGCGCGCGGCTGCGACCTGACTGTGCAATACCGTACATACCCAGCGGATACG 420
DB 542 GCTTGTGCGCGCGGCTGCGACCTGACTGTGCAATACCGTACATACCCAGCGGATACG 601
QY 421 TTCAATGCGGATCCGTTATTTAAACCTGTAACAACTGCGGTTCCCACTGGATACGCG 480
DB 602 TTCAATGCGGATCCGTTATTTAAACCTGTAACAACTGCGGTTCCCACTGGATACGCG 661
QY 481 AACGTGATGACGCGATCTCAGACGAGGCAAGAGGTCATTTGCTGACTTTACCGGCAT 540
DB 662 AACGTGATGACGCGATCTCAGACGAGGCAAGAGGTCATTTGCTGACTTTACCGGCAT 721
QY 541 CGGCAACCGCGGTTTCCGGAACCTGAAACGGGTGCTTAAATTTTCCGCAATCAAATTGTC 600
DB 722 CGGCAACCGCGGTTTCCGGAACCTGAAACGGGTGCTTAAATTTTCCGCAATCAAATTGTC 781
QY 601 CTTAAAGCTGAGAAACGAGAAAGCTGTTCAATTAACGAGGATTAACCATCGGAATC 660
DB 782 CTTAAAGCTGAGAAACGAGAAAGCTGTTCAATTAACGAGGATTAACCATCGGAATC 841
QY 661 AAGGTGAGCGCGCAATGCTCATTAACCGGTGCGGTAAAGCTCTGCAATCATGCTGACG 720
DB 842 AAGGTGAGCGCGCAATGCTCATTAACCGGTGCGGTAAAGCTCTGCAATCATGCTGACG 901
QY 721 GAGATATTTTCTCTGCAACAAACAGGGAATGCCGAGCGGAGTGGGGAAGATCAC 780

DB 902 GAAATATTTTCTCTGCAACAAACAGGAAATGCCGAGCCGCGGAGGGAAGATCACT 961
QY 781 GATTCACACCGTGGAACACTTGTCTAATGTTGCAATTAAGCCCAATTTATTTGCTACAA 840
DB 962 GATTCACACCGTGGAACACTTGTCTAATGTTGCAATTAAGCCCAATTTATTTACTACAA 1021
QY 841 CGCAGCGCAGAGTTGCCCGGAGCGCGCACCCCGTATTTAGATTGATCAAGACAGCG 900
DB 1022 CGCAGCGCAGAGTTGCCCGGAGCGCGCACCCCGTATTTAGATTGATCAAGACAGCG 1081
QY 901 TTGACCGCCCATCAACCGCAAAACAGCGGTATGTTGATCAATTAACCACTTCAGTCTG 960
DB 1082 TTGACCGCCCATCAACCGCAAAACAGCGGTATGTTGATCAATTAACCACTTCAGTCTG 1141
QY 961 TTTATGCGCGGACAGATTAATCTATCTGCAATCTTGGGCGGCGCATCTGAGCTCACTG 1020
DB 1142 TTTATGCGCGGACAGATTAATCTATCTGCAATCTTGGGCGGCGCATCTGAGCTCACTG 1201
QY 1021 ACCCTTCCCGGTGAGCGGATTAACAGCGCGCAGGTGTAATCTGTTGAAACGCTGG 1080
DB 1202 ACCCTTCCCGGTGAGCGGATTAACAGCGCGCAGGTGTAATCTGTTGAAACGCTGG 1261
QY 1081 CGTGGCTAAGCGATTAACAGCGCATGATTAAGTTCGCTGTCTTCCAGACTTTACAG 1140
DB 1262 CGTGGCTAAGCGATTAACAGCGCATGATTAAGTTCGCTGTCTTCCAGACTTTACAG 1321
QY 1141 CAGATCGGTATTAACCGCGCTGTCAATTAATACCGCGCGCGAAGGTGAACCTGACC 1200
DB 1322 CAGATCGGTATTAACCGCGCTGTCAATTAATACCGCGCGCGAAGGTGAACCTGACC 1381
QY 1201 CTGGCAGATGTAAGCAAGCAATGCGCAGGCGATGTTGTTGGCAGGTTTACGCA 1260
DB 1382 CTGGCAGATGTAAGCAAGCAATGCGCAGGCGATGTTGTTGGCAGGTTTACGCA 1441
QY 1261 ATCGTAATGAACAGCATACCCGCTTGGCAATTTGAA 1299
DB 1442 ATCGTAATGAACAGCATACCCGCTTGGCAATTTGAA 1480

RESULT 6
US-08-910-798-1
Sequence 1, Application US/08910798
Patent No. 5876997
GENERAL INFORMATION:
APPLICANT: KREIZ
TITLE OF INVENTION: NOVEL PHYTASE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: California
COUNTRY: US
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/910,798
FILING DATE: August 13, 1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: HAILE, PH.D., LISA A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/029001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5099
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:

```

;
; LENGTH: 1272 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; IMMEDIATE SOURCE:
; CLONE: PHYTASE
; FEATURE:
; NAME/KEY:
; LOCATION:
;
US-08-910-798-1

Query Match      90.1%; Score 1170.2; DB 2; Length 1272;
Best Local Similarity 95.4%; Pred. No. 0;
Matches 1236; Conservative 0; Mismatches 9; Indels 51; Gaps 1;

QY 1 ATGAAAGCCATCTTATCCCATTTTATCTCTTGATTCGCTTAAACCCGCAATCTCA 60
DB 1 ATGAAAGCCATCTTATCCCATTTTATCTCTTGATTCGCTTAAACCCGCAATCTCA 60
QY 61 TTGCTCAGAGTAGCCGAGCTGAAGCTGAAAGTGTGATGTCAGTGTGATGTT 120
DB 61 TTGCTCAGAGTAGCCGAGCTGAAGCTGAAAGTGTGATGTCAGTGTGATGTT 120
QY 121 GTGCGTCTCAACCAAGCCACGCACTGATGACAGATGTCAACCCGACAGCGATGGCA 180
DB 121 GTGCGTCTCAACCAAGCCACGCACTGATGACAGATGTCAACCCGACAGCGATGGCA 180
QY 181 ACCGTGCGCGTAAACACTGGGTTGGCTGACACCGCGCGGTGTGAGCTAATCCCTATCTC 240
DB 181 ACCGTGCGCGTAAACACTGGGTTGGCTGACACCGCGCGGTGTGAGCTAATCCCTATCTC 240
QY 241 GGAACATTAACCAAGCCACGCTGTAGCCGAGATTGCTGGCAAAAAGGCTCCCG 300
DB 241 GGAACATTAACCAAGCCACGCTGTAGCCGAGATTGCTGGCAAAAAGGCTCCCG 300
QY 301 CAGTCTGCTCAAGTCCGATTTATTCGTGATGTCAGAGGTCACCGGTAAACAGCGAA 360
DB 301 CAGTCTGCTCAAGTCCGATTTATTCGTGATGTCAGAGGTCACCGGTAAACAGCGAA 360
QY 361 GCCTTGGCGCGCGGGCTGGCACTGACTGTGCAATACCGTAATACCCAGGAGATACG 420
DB 361 GCCTTGGCGCGCGGGCTGGCACTGACTGTGCAATACCGTAATACCCAGGAGATACG 420
QY 357 -----CCAGGACGATACG 369
DB 357 -----CCAGGACGATACG 369
QY 421 TCCAGTCCCGATCCGTTATTTATTCCTCTTAAAACTGGCGTTGCCAATCGAATAACGG 480
DB 421 TCCAGTCCCGATCCGTTATTTATTCCTCTTAAAACTGGCGTTGCCAATCGAATAACGG 480
QY 481 AAAGTGAAGTGAAGGATCTGACGAGGAGGAGGTCATTTGCTGACTTTAACCGGAGAT 540
DB 481 AAAGTGAAGTGAAGGATCTGACGAGGAGGAGGTCATTTGCTGACTTTAACCGGAGAT 540
QY 541 CGGCAAAAGCGGTTTCGGAACGTGAACGGGTGCTTAATTTTCCGCAATCAAACTTTGTC 600
DB 541 CGGCAAAAGCGGTTTCGGAACGTGAACGGGTGCTTAATTTTCCGCAATCAAACTTTGTC 600
QY 601 CTTAAAGCTGAGAAACAGAGAGAAAGCTGTCATTAAAGCAGGATTAACATCGGAATC 660
DB 601 CTTAAAGCTGAGAAACAGAGAGAAAGCTGTCATTAAAGCAGGATTAACATCGGAATC 660
QY 661 AAGGTAGAGCGCGCAATGTCTCATTAACCGGTGCGTAACTTCGATCAATGCTGACG 720
DB 661 AAGGTAGAGCGCGCAATGTCTCATTAACCGGTGCGTAACTTCGATCAATGCTGACG 720
QY 721 GAGATATTTCTCTGCAACAGACAGGAGATGCGGAGCGGGGTGGGAAAGATCAAC 780
DB 721 GAGATATTTCTCTGCAACAGACAGGAGATGCGGAGCGGGGTGGGAAAGATCAAC 780
QY 781 GATTCAACACAGTGAACCTTTGCTAGTTTGAATACGCGCAATTTTATTTGCTACAA 840
DB 781 GATTCAACACAGTGAACCTTTGCTAGTTTGAATACGCGCAATTTTATTTGCTACAA 840
QY 841 CGCAGCGCAGAGGTTGCCGACGCGCCACCCCGTTATTGATTTGATCAAGACGCG 900
DB 841 CGCAGCGCAGAGGTTGCCGACGCGCCACCCCGTTATTGATTTGATCAAGACGCG 900

```

```

DB 790 CGCAGCCGAGAGTTGCCCGACGCGCCGACCCCGTTATTGATTTGATCAATGACAGG 849
QY 901 TTGACGCCCATTCACCCGAAAAACAGGGGTATNGGTGACATTAACCCACTTCAGTGTG 960
DB 901 TTGACGCCCATTCACCCGAAAAACAGGGGTATNGGTGACATTAACCCACTTCAGTGTG 960
QY 961 TTTATGCGCGGACAGCATACTAATCTGGCAAACTTCGCGGCGCATGAGACTCAACTGG 1020
DB 961 TTTATGCGCGGACAGCATACTAATCTGGCAAACTTCGCGGCGCATGAGACTCAACTGG 1020
QY 1021 ACCGTTCCCGGTGACCGCGATTAACAGCCGCGCCAGGTGTGAACCTGGTTTGAAGCTGG 1080
DB 1021 ACCGTTCCCGGTGACCGCGATTAACAGCCGCGCCAGGTGTGAACCTGGTTTGAAGCTGG 1080
QY 970 ACGCTTCCCGGTGACCGCGATTAACAGCCGCGCGGTGTGAACCTGGTTTGAAGCTGG 1029
DB 970 ACGCTTCCCGGTGACCGCGATTAACAGCCGCGCGGTGTGAACCTGGTTTGAAGCTGG 1029
QY 1081 CGTGCGTAAAGCATTAACAGCCAGTGTGATTCAGTTTCGCTGCTTCCGACATTTACAG 1140
DB 1081 CGTGCGTAAAGCATTAACAGCCAGTGTGATTCAGTTTCGCTGCTTCCGACATTTACAG 1140
QY 1141 CAGATGCGTAAAGCATTAACAGCCAGTGTGATTCAGTTTCGCTGCTTCCGACATTTACAG 1200
DB 1141 CAGATGCGTAAAGCATTAACAGCCAGTGTGATTCAGTTTCGCTGCTTCCGACATTTACAG 1200
QY 1201 CTGCGAGAGTGTGAAGCCGAATGCGCAGAGGATGTTGTTGGCAGGTTTACGCA 1260
DB 1201 CTGCGAGAGTGTGAAGCCGAATGCGCAGAGGATGTTGTTGGCAGGTTTACGCA 1260
QY 1261 ATCGGAATGAAGCAGCATACCGCTTGCAGTTTG 1296
DB 1261 ATCGGAATGAAGCAGCATACCGCTTGCAGTTTG 1296
QY 1210 ATCGGAATGAAGCAGCATACCGCTTGCAGTTTG 1245
DB 1210 ATCGGAATGAAGCAGCATACCGCTTGCAGTTTG 1245

```

RESULT 7

```

US-09-489-039A-341
; Sequence 341, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 341
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
; US-09-489-039A-341

```

```

Query Match      4.8%; Score 63; DB 4; Length 1266;
Best Local Similarity 48.7%; Pred. No. 2e-10;
Matches 202; Conservative 0; Mismatches 210; Indels 3; Gaps 1;

QY 86 AGCTGAAAGTGTGATTTGATTCAGTCTGATGCTGTGGTGTCTCC---AACCAAGGCA 142
DB 113 AGCTGAGAGGTTGCTGATCATGAGCCGTCAACAACCTGTGTGCGCTTTGGCAATACG 172
QY 143 CGCAACTGATGAGATGTCACCCGACGATGCAACTGCGCGTTAAACTGGGTT 202
DB 173 GCAAGCTGTGAAACAGTCAACCGCCAGAGGCTGTGGCGGAGTGTGCGGCGGCGC 232
QY 203 GGTGACACCGCGCGGTGTGAGCTAATGCTGATCTCGGACATTAACAGCCAGCTTC 262
DB 233 AGCTGACACCAAAAGCGGTGTGCTGAGAGTGTATGTGGCCACTATATGCGGATGGC 292
QY 263 TGTGACCGACGAGTGTGCTGTGGGAAAAAGGCTGCGCGAGTGTGAGTTCGAGATTA 322
DB 293 TGGCGCAGCAGAAAGTGTGTGACCAAGCGGAGAGTGTGCGCGGAAAAAGGCTTTAAGCCT 352
QY 323 TTGCTGATGTGACAGAGCTTACCGTTAAACAGGCGAAGCCTTTCGCGCGGCGTGGCAC 382
DB 323 TTGCTGATGTGACAGAGCTTACCGTTAAACAGGCGAAGCCTTTCGCGCGGCGTGGCAC 382

```

```

Db      353  ACCGTTACACGCTCCGACGCCACCGTCCGACCGCGAGTTCTTTATCATCCGCGCGCTTC 412
Qy      383  CTGACTGTGCAATTAACCGTACATACCCAGGACAGATACGTCCAGTCCGATTATTA 442
Db      413  CGGGCTGGGGGATCCGGGTGATCACAGCTCAGATGGGACATGAGACCGGACCTTCA 472
Qy      443  ATCCCTTAAAACTGGCGTTTGGCAACTGATTAACCGGAACGTGACTGAGCGCAT 497
Db      473  ACCCGGTTATTAACCGACGACTCCCGCGCTTTGCGAGAAAGCGCTGACGCGCAT 527

```

RESULT 8

```

US-09-710-794-4/c
; Sequence 4, Application US/09710794
; Patent No. 6573069
; GENERAL INFORMATION:
; APPLICANT: Holloway, James L.
; APPLICANT: Gao, Zeren
; APPLICANT: Whitmore, Theodore E.
; TITLE OF INVENTION: NOVEL CRIB PROTEIN ZMS1
; FILE REFERENCE: 99-76
; CURRENT APPLICATION NUMBER: US/09/710,794
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/164,685
; PRIOR FILING DATE: 1999-11-10
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 2868
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (173)...(1219)
US-09-710-794-4

```

```

Query Match      3.1%; Score 40.8; DB 4; Length 2868;
Best Local Similarity 53.8%; Pred. No. 0.012;
Matches 84; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

```

```

Qy      128  CTCGAACCAAGCGCCAGCACTGATGAGATGTCACCCAGACGATGGCCAACTGGC 187
Db      353  CTCCTCTCCCTGGCTTGCTGTAGAGAGAGGTGTCCTCCAGCGCTCCAGCCGGC 234
Qy      188  CGGTAAAACTGGTGGCTGACACCGCGCGGTGAGTAACTGCTATCTCGACATT 247
Db      293  CCACATGATGCTGTGTGCGGAGATGACCCAGCGGGGCACTGATCATCTCGGTGAGT 234
Qy      248  ACCAAGCGCAGCGCTTGTGAGCCGACGAGATTGCTGG 283
Db      233  CCGCAGGTGAGCGGCGCTTGAATTCAAGAGCTGG 198

```

RESULT 9

```

US-09-252-991A-15917
; Sequence 15917, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 15917
; LENGTH: 480
; TYPE: DNA

```

```

; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15917

```

```

Query Match      2.8%; Score 36.4; DB 4; Length 480;
Best Local Similarity 55.6%; Pred. No. 0.12;
Matches 70; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

```

```

Qy      254  GCCAGGCTGTGTAGCCGACGATTTGCTGCGAATAAGGGCTGCCGACGTGTCAG 313
Db      155  GCCAGGCTGTGCGCGCGGACATCAAGCGGGCGATTCAGCCGGGCTGATCAG 214
Qy      314  TCGCATTAATGCTGATGTGACGAGCGTACCCGTAAACAGCGGAGCTTGCCTCG 373
Db      215  TAGCGGTAGGGCTGTGTGCGGACGAGCGCGGAAATCCGCGACAGGCTTCAGCTTG 274
Qy      374  GGCTGG 379
Db      275  CGCTGG 280

```

RESULT 10

```

US-09-252-991A-15888
; Sequence 15888, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 15888
; LENGTH: 762
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15888

```

```

Query Match      2.8%; Score 36.4; DB 4; Length 762;
Best Local Similarity 55.6%; Pred. No. 0.16; 56; Indels 0; Gaps 0;
Matches 70; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

```

```

Qy      254  GCCAGGCTGTGTAGCCGACGATTTGCTGCGAATAAGGGCTGCCGACGTGTCAG 313
Db      24  GCCAGGCTGTGCGCGCGGACATCAAGCGGGCGATTCAGCCGGGCTGATCAG 83
Qy      314  TCGCATTAATGCTGATGTGACGAGCGTACCCGTAAACAGCGGAGCTTGCCTCG 373
Db      84  TAGCGGTAGGGCTGTGTGCGGACGAGCGCGGAAATCCGCGACAGGCTTCAGCTTG 143
Qy      374  GGCTGG 379
Db      144  CGCTGG 149

```

RESULT 11

```

US-09-252-991A-15774/c
; Sequence 15774, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190

```

PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 15774
 LENGTH: 885
 TYPE: DNA
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-15774

Query Match 2.8%; Score 36.4; DB 4; Length 885;
 Best Local Similarity 55.6%; Pred. No. 0.18;
 Matches 70; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 254 GCCAGCGCTGTGTAGCCGACGATGCTGCGAAAAAGGCGTCCCGCAGTGTGTCAGG 313
 DB 735 GCCAGCGCTGTGCGCCGCGACATCAGCCGCGCGCATGTCCAGCCGCGCTGTGATCAGG 676
 QY 314 TCGCGATTATTGTCTATGTTCAGACGACCTACCCGTAAACAGCGGAGCCTTCGCGCGG 373
 DB 675 TAGCGGTAGGGGCTGTGTGCGCAACAGCGCGGAAATCCGCGACAGGCTCCAGGCTTGG 616
 QY 374 GGCTGG 379
 DB 615 CGCTGG 610

RESULT 12

US-09-252-991A-15804/C
 Sequence 15804, Application US/09252991A
 Patent No. 6551795
 GENERAL INFORMATION:
 APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 CURRENT FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 15804
 LENGTH: 1068
 TYPE: DNA
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-15804

Query Match 2.8%; Score 36.4; DB 4; Length 1068;
 Best Local Similarity 55.6%; Pred. No. 0.2;
 Matches 70; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 254 GCCAGCGCTGTGTAGCCGACGATGCTGCGAAAAAGGCGTCCCGCAGTGTGTCAGG 313
 DB 956 GCCAGCGCTGTGCGCCGCGACATCAGCCGCGCGCATGTCCAGCCGCGCTGTGATCAGG 897
 QY 314 TCGCGATTATTGTCTATGTTCAGACGACCTACCCGTAAACAGCGGAGCCTTCGCGCGG 373
 DB 896 TAGCGGTAGGGGCTGTGTGCGCAACAGCGCGGAAATCCGCGACAGGCTCCAGGCTTGG 837
 QY 374 GGCTGG 379
 DB 836 CGCTGG 831

RESULT 13

US-09-710-794-1/C
 Sequence 1, Application US/09710794
 Patent No. 6573069
 GENERAL INFORMATION:
 APPLICANT: Holloway, James L.
 APPLICANT: Gao, Zeren
 APPLICANT: Whitmore, Theodore F.
 TITLE OF INVENTION: NOVEL CRIB PROTEIN ZMSEI

FILE REFERENCE: 99-76
 CURRENT APPLICATION NUMBER: US/09/710,794
 CURRENT FILING DATE: 2000-11-09
 PRIOR APPLICATION NUMBER: US 60/164,685
 PRIOR FILING DATE: 1999-11-10
 NUMBER OF SEQ ID NOS: 31
 SOFTWARE: FaastSeq for Windows Version 3.0
 SEQ ID NO 1
 LENGTH: 3076
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (199)...(1266)
 US-09-710-794-1

Query Match 2.7%; Score 35.4; DB 4; Length 3076;
 Best Local Similarity 51.6%; Pred. No. 0.89;
 Matches 81; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 139 GCCACGAACCTGATGACGAGATGTCAACCCAGACGATGACCAACTGCGGTAACCTG 198
 DB 368 GCTTGCTATTGAGAGAGAGGTGTCCCAAGGCGTCTCGGCGCGCAAGTGCATG 309
 QY 199 GATTGCTGACACCGCGGTGTGAGCTAATCGCTTATCTGGAATTACCAAGCCAG 258
 DB 308 GTGTGGCGGAGAGTGGCCAGCGGGCGCTGATCATCTCGCGCTGAGAGTCCGCTGGGAA 249
 QY 259 GGTCTGTAGCCGACGAGATTGCTGCGCAAAAGGCT 295
 DB 248 CGGCGCTTGAGTGCACGAGCTGACACCACTTGTCT 212

RESULT 14

US-09-902-540-2830
 Sequence 2830, Application US/09902540
 Patent No. 6833447
 GENERAL INFORMATION:
 APPLICANT: Goldman, Barry S.
 APPLICANT: Goldman, Hinkle, Gregory J.
 APPLICANT: Slater, Steven C.
 APPLICANT: Wiegand, Roger C.
 TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
 FILE REFERENCE: 38-10(15849)B
 CURRENT APPLICATION NUMBER: US/09/902,540
 CURRENT FILING DATE: 2001-07-10
 PRIOR APPLICATION NUMBER: 60/217,883
 PRIOR FILING DATE: 2000-07-10
 NUMBER OF SEQ ID NOS: 16825
 SEQ ID NO 2830
 LENGTH: 2232
 TYPE: DNA
 ORGANISM: Myxococcus xanthus
 US-09-902-540-2830

Query Match 2.7%; Score 34.6; DB 4; Length 2232;
 Best Local Similarity 56.6%; Pred. No. 1.3;
 Matches 64; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 268 GCCGACGATTGCTGTGCGAAAAAGGCGTCCCGCAGTGTGTCAGTGTGCGATTATTGCT 327
 DB 1294 GCCGCCCAATCTGTGAGAGACCCGGGCGCGCTCGTCCGCTCGGCGCGAGCATGGCG 1353
 QY 328 GATGTGACAGAGGTACCCGTAAACAGGCGAAGCCTTGCCTCGCGGCTGGC 380
 DB 1354 GCGGTGATGATGAGGACCCGAGATGAGAGACGAGAGCTGACGACGAGAGGAC 1406

RESULT 15

US-09-902-540-1012/C
 Sequence 1012, Application US/09902540
 Patent No. 6833447
 GENERAL INFORMATION:

APPLICANT: Goldman, Barry S.
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Wiegand, Roger C.
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
FILE REFERENCE: 38-10(15849)B
CURRENT APPLICATION NUMBER: US/09/902,540
CURRENT FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: 60/217,883
PRIOR FILING DATE: 2000-07-10
NUMBER OF SEQ ID NOS: 16825
SEQ ID NO 1012
LENGTH: 14570
TYPE: DNA
ORGANISM: Myxococcus xanthus
US-09-902-540-1012

Query Match 2.7%; Score 34.6; DB 4; Length 14570;
Best Local Similarity 56.6%; Pred. No. 4.5;
Matches 64; Conservative 0; Mismatches 49; Indels 0; Gaps 0;
QY 268 GCCGACGATGCTGCTGGCGAAGAGGCTGCCGAGTGTGTCAGTTCGATTTATTTGCT 327
DB 10464 GCCGCCCAATCTGTGAGAGACCCGGGGCCGCTCCGTCGCGCGGAGCATGGCG 10405
QY 328 GATGTCGACGAGCGTAACCGGTAACAGCGGAAGCTTCGCCCGGGGCTGGC 380
DB 10404 GCGGTCGATGAGCGCACCCGATGAGAGCAGAGAGCTGGACGACGACGAC 10352

Search completed: February 10, 2005, 17:43:01
Job time : 276.446 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 22:13:58 ; Search time 946.54 Seconds
(without alignments)
8097.594 Million cell updates/sec

Title: US-09-926-375b-7_COPY_12653_13951

Perfect score: 1299
Sequence: 1 atgaagaccattatccccc.....taccgcgttcagttgttaa 1299

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 5378673 seqs, 295022984 residues

Total number of hits satisfying chosen parameters: 10757346

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/prodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/prodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/prodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/prodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/prodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/prodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/prodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/prodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/prodata/2/pubpna/US09C_PUBCOMB.seq:*
11: /cgn2_6/prodata/2/pubpna/US09C_NEW_PUB.seq:*
12: /cgn2_6/prodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/prodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/prodata/2/pubpna/US10C_PUBCOMB.seq:*
15: /cgn2_6/prodata/2/pubpna/US10E_PUBCOMB.seq:*
16: /cgn2_6/prodata/2/pubpna/US10F_PUBCOMB.seq:*
17: /cgn2_6/prodata/2/pubpna/US10F_PUBCOMB.seq:*
18: /cgn2_6/prodata/2/pubpna/US10F_PUBCOMB.seq:*
19: /cgn2_6/prodata/2/pubpna/US11_NEW_PUB.seq:*
20: /cgn2_6/prodata/2/pubpna/US11_NEW_PUB.seq:*
21: /cgn2_6/prodata/2/pubpna/US60_NEW_PUB.seq:*
22: /cgn2_6/prodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1294.2	99.6	1299	17	US-10-282-122A-7167
2	1293.2	99.6	1901	9	US-09-866-379-7
3	1293.2	99.6	1901	15	US-10-156-660-3
4	1293.2	99.6	1901	17	US-10-601-319-7
5	1291.6	99.4	1901	9	US-09-866-379-5
6	1291.6	99.4	1901	9	US-09-866-379-9
7	1290	99.3	1901	17	US-10-601-319-5
8	1288.4	99.2	1901	9	US-09-866-379-6
9	1288.4	99.2	1901	17	US-10-601-319-6
10	1282.2	98.7	1323	9	US-09-777-566A-1
11	1282.2	98.7	1323	9	US-09-866-379-1

12	1282.2	98.7	1323	14	US-10-034-985-1
13	1282.2	98.7	1323	17	US-10-430-356-1
14	1282.2	98.7	1323	17	US-10-601-319-1
15	1259.2	96.9	1308	15	US-10-156-660-1
16	1259.2	96.9	1308	17	US-10-601-319-9
17	1257.4	96.8	1489	14	US-10-266-041-9
18	1257.4	96.8	1489	17	US-10-284-962-1
19	1252.6	96.4	1486	17	US-10-284-962-4
20	658	51.4	1281	16	US-10-334-672-4
21	658	51.4	1281	16	US-10-334-671-4
22	403.8	31.1	466	17	US-10-282-122A-3181
23	228.6	17.6	1326	17	US-10-282-122A-41608
24	224.6	17.3	1325	15	US-10-021-723A-3
25	221.4	17.0	1325	15	US-10-021-723A-1
26	210.8	16.2	1326	15	US-10-021-723A-11
27	162.6	12.5	254	17	US-10-317-444-301
28	162.6	12.5	254	17	US-10-317-444-302
29	162.6	12.5	254	17	US-10-317-444-303
30	162.6	12.5	254	17	US-10-317-444-304
31	157.4	12.1	1431	15	US-10-021-723A-9
32	74	5.7	1230	15	US-10-021-723A-7
33	65.8	5.1	1266	15	US-10-021-723A-5
34	56.2	4.3	11710	16	US-10-240-689-40
35	39.4	3.0	2256646	18	US-10-470-565-1
36	39	3.0	1068	18	US-10-437-963-53530
37	38	2.9	1395	17	US-10-369-493-41471
38	37.2	2.9	2879	17	US-10-108-260A-2180
39	36	2.8	450	18	US-10-357-930-5437
40	35.6	2.7	427	17	US-10-393-840-372
41	35.4	2.7	344	18	US-10-357-930-4929
42	35.4	2.7	349	18	US-10-357-930-14098
43	35.4	2.7	391	18	US-10-357-930-35227
44	35.4	2.7	391	18	US-10-357-930-44068
45	35.4	2.7	405	18	US-10-357-930-14606

ALIGNMENTS

RESULT 1
US-10-282-122A-7167
; Sequence 7167, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zykkind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: EPIRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625

Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 9, Appli
Sequence 9, Appli
Sequence 9, Appli
Sequence 1, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 3181, Ap
Sequence 41608, A
Sequence 3, Appli
Sequence 11, Appli
Sequence 301, App
Sequence 302, App
Sequence 303, App
Sequence 304, App
Sequence 9, Appli
Sequence 7, Appli
Sequence 5, Appli
Sequence 40, Appli
Sequence 53530, A
Sequence 41471, A
Sequence 2180, Ap
Sequence 5437, Ap
Sequence 372, App
Sequence 4929, Ap
Sequence 14098, A
Sequence 35227, A
Sequence 44068, A
Sequence 14606, A

```

; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See file Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7167
; LENGTH: 1299
; TYPE: DNA
; ORGANISM: Escherichia coli
; US-10-282-122A-7167

Query Match      99.6%; Score 1294.2; DB 17; Length 1299;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1296; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 ATGAAAGCCATCTTAATCCCATTTTATCTTCTGATTCGGTAAACCCCGCAATCTGCA 60
DB      1 ATGAAAGCCATCTTAATCCCATTTTATCTTCTGATTCGGTAAACCCCGCAATCTGCA 60
QY      61 TTGCGTCAGATGAGCGCGGAGCTGAAGCTGAAGTGTGATTTGTCAGTGTGATGGT 120
DB      61 TTGCGTCAGATGAGCGCGGAGCTGAAGCTGAAGTGTGATTTGTCAGTGTGATGGT 120
QY      121 GTGCGTCTCAACCAAGGCGCAAGCACTGATGAGATGTCAACCCGACGATGGCCA 180
DB      121 GTGCGTCTCAACCAAGGCGCAAGCACTGATGAGATGTCAACCCGACGATGGCCA 180
QY      122 GTGCGTCTCAACCAAGGCGCAAGCACTGATGAGATGTCAACCCGACGATGGCCA 180
DB      122 GTGCGTCTCAACCAAGGCGCAAGCACTGATGAGATGTCAACCCGACGATGGCCA 180
QY      181 ACCGCGCGGTAAACTGGGTGGTGTGACACCGCGGGGTGTGAGTAACTCCCTATCTC 240
DB      181 ACCGCGCGGTAAACTGGGTGGTGTGACACCGCGGGGTGTGAGTAACTCCCTATCTC 240
QY      181 ACCGCGCGGTAAACTGGGTGGTGTGACACCGCGGGGTGTGAGTAACTCCCTATCTC 240
DB      181 ACCGCGCGGTAAACTGGGTGGTGTGACACCGCGGGGTGTGAGTAACTCCCTATCTC 240
QY      241 GGACATTAACCAAGCGAGCGGTCTGTGAGCGCAAGATTCCTGACGAAAAAGGAGCTGCCG 300
DB      241 GGACATTAACCAAGCGAGCGGTCTGTGAGCGCAAGATTCCTGACGAAAAAGGAGCTGCCG 300
QY      301 CAGTCTGATCAGGTCCGATTAATGCTGATGTCGACGACGTAACCCGTAACCAAGCGAA 360
DB      301 CAGTCTGATCAGGTCCGATTAATGCTGATGTCGACGACGTAACCCGTAACCAAGCGAA 360
QY      361 GCGTTCGCGCGCGGCGTGGCACTGCTGTGCAATACCGTACATACCCAGGCAAGTACG 420
DB      361 GCGTTCGCGCGCGGCGTGGCACTGCTGTGCAATACCGTACATACCCAGGCAAGTACG 420
QY      421 TCCAGTCCCGATCCGTTATTTAATCTCTAAATACTGGGTTTGCACATGATTAACGCG 480
DB      421 TCCAGTCCCGATCCGTTATTTAATCTCTAAATACTGGGTTTGCACATGATTAACGCG 480
QY      481 AACGTGATCGACGCGATCTCTACAGAGGGAAGAGGGTCAATTGCTGACTTTAACCGGGCAT 540
DB      481 AACGTGATCGACGCGATCTCTACAGAGGGAAGAGGGTCAATTGCTGACTTTAACCGGGCAT 540
QY      541 CCGCAACAGCGCTTTGGCGCACTGGAACCGGCTTAATTTTCCGAAATCAAACTTTGTCG 600
DB      541 CCGCAACAGCGCTTTGGCGCACTGGAACCGGCTTAATTTTCCGAAATCAAACTTTGTCG 600
QY      601 CTTAAACGTGAGAAACAGACGAAACCTGTTCACTTAACGACGATTAACATGGAATCTC 660
DB      601 CTTAAACGTGAGAAACAGACGAAACCTGTTCACTTAACGACGATTAACATGGAATCTC 660
QY      661 AAGGTAGCGCGCAATGCTCATTTAACCGGTGCGGTAAAGCTGCAATCAATGCTGACG 720
DB      661 AAGGTAGCGCGCAATGCTCATTTAACCGGTGCGGTAAAGCTGCAATCAATGCTGACG 720
QY      721 GAGATTTTCTCTCTGCAACAGACAGGAAATGCCGAGCCGGGGTGGGGAGAGATCAC 780
DB      721 GAGATTTTCTCTCTGCAACAGACAGGAAATGCCGAGCCGGGGTGGGGAGAGATCAC 780
QY      781 GATTACACACGAGGAACCTTGTAAGTTTGCAATACGCGCAATTTATTGTCTACA 840
DB      781 GATTACACACGAGGAACCTTGTAAGTTTGCAATACGCGCAATTTATTGTCTACA 840
```

```

DB      781 GATTACACACGAGGAACCTTGTAAGTTTGCAATACGCGCAATTTATTGTCTACAA 840
QY      841 CGCACGCCAGAGGTTGCCCGGACCGCGCCACCCCGTTATTAAGTTTGAATAACAGCG 900
DB      841 CGCACGCCAGAGGTTGCCCGGACCGCGCCACCCCGTTATTAAGTTTGAATAACAGCG 900
QY      901 TTGACGCCCATCCACCGCAAAAACAGCGATGATGTGATGACATTAACCACTTCACTGCTG 960
DB      901 TTGACGCCCATCCACCGCAAAAACAGCGATGATGTGATGACATTAACCACTTCACTGCTG 960
QY      961 TTTATCGCCGACACGATTAATCTGCAAAATCTCGCGGCGCACTGAGCTCAACTGG 1020
DB      961 TTTATCGCCGACACGATTAATCTGCAAAATCTCGCGGCGCACTGAGCTCAACTGG 1020
QY      1021 ACGTTCCCGGTCAAGCCGGAATAACACCGCCGAGGTGTGAATCTGTGTTGAACGCTGG 1080
DB      1021 ACGTTCCCGGTCAAGCCGGAATAACACCGCCGAGGTGTGAATCTGTGTTGAACGCTGG 1080
QY      1081 CGTGGCTTAAGCGATTAACGACGAGTGAATTCAGGTTTGTGCTGTGCTTCCAGACTTTACAG 1140
DB      1081 CGTGGCTTAAGCGATTAACGACGAGTGAATTCAGGTTTGTGCTGTGCTTCCAGACTTTACAG 1140
QY      1141 CAGATCGGTGAATAAACCGCGCTGTCAATTAATAACCGCCCGGAGAGGTGAACCTGACC 1200
DB      1141 CAGATCGGTGAATAAACCGCGCTGTCAATTAATAACCGCCCGGAGAGGTGAACCTGACC 1200
QY      1201 CTGGCAGATGTGAAGACGAAATGCGAGGCGATGTGTTGTTGGCGAGTTTAAACGCA 1260
DB      1201 CTGGCAGATGTGAAGACGAAATGCGAGGCGATGTGTTGTTGGCGAGTTTAAACGCA 1260
QY      1261 ATCGTAATGAAGACGATACCGCTTGCAGTTTGTAA 1299
DB      1261 ATCGTAATGAAGACGATACCGCTTGCAGTTTGTAA 1299

RESULT 2
US-09-866-379-7
; Sequence 7, Application US/09866379
; Patent No. US20020136754A1
; GENERAL INFORMATION:
; APPLICANT: DIVERSA CORPORATION
; APPLICANT: SHORT, Jay
; APPLICANT: KRETT, Keith
; APPLICANT: GRAY, Kevin
; APPLICANT: BARTON, Nelson
; APPLICANT: GARRETT, James
; APPLICANT: O'DONOGHUE, Eileen
; TITLE OF INVENTION: RECOMBINANT BACTERIAL PHYTASES AND USES THEREOF
; FILE REFERENCE: DIVER1370-7
; CURRENT APPLICATION NUMBER: US/09/866,379
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: US 09/580,515
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 09/318,528
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: US 09/291,931
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: US 09/259,214
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: US 08/910,798
; PRIOR FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 1901
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(1901)
; OTHER INFORMATION: n is any nucleotide
; US-09-866-379-7
```

Query Match 99.6%; Score 1293.2; DB 9; Length 1901;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1295; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

1 ATGAAAGCCATCTTAATCCATTTTATCTCTTGTATCCGTAAACCCCGAATCTGCA 60
188 ATGAAAGCCATCTTAATCCATTTTATCTCTTGTATCCGTAAACCCCGAATCTGCA 247
61 TTGCGCTCAGAGTGAAGCCGAGCTGAAGTGAAGTGTGATGTCTAGTCTCATGTG 120
248 TTGCGCTCAGAGTGAAGCCGAGCTGAAGTGAAGTGTGATGTCTAGTCTCATGTG 307
121 GTGCGTCTCAACCAAGGCCAAGCACTGATGACAGATGTCAACCCAGACGATGCGCA 180
308 GTGCGTCTCAACCAAGGCCAAGCACTGATGACAGATGTCAACCCAGACGATGCGCA 367
181 ACCTGCGCGGTAAACTGTGGTGTGATGACACCGCGGTGTGAGTAAATCGCTATCTC 240
368 ACCTGCGCGGTAAACTGTGGTGTGATGACACCGCGGTGTGAGTAAATCGCTATCTC 427
241 GGACATTAACCAAGCCGAGCTGTAGCCGACGATGCTGCGAAGGAGGCTGCGCG 300
428 GGACATTAACCAAGCCGAGCTGTAGCCGACGATGCTGCGAAGGAGGCTGCGCG 487
301 CAGCTGTGATCAGGTGCGCATTTATGTGATGTGACAGCGGTAACCCGTAAACAGCGCA 360
488 CAGCTGTGATCAGGTGCGCATTTATGTGATGTGACAGCGGTAACCCGTAAACAGCGCA 547
361 GCGTTTGGCGCGCGGCTGCGCACTGACTGTGCAATACCTAATACCCAGGAGATACG 420
548 GCGTTTGGCGCGCGGCTGCGCACTGACTGTGCAATACCTAATACCCAGGAGATACG 607
421 TCCAGTCCCGATCGTTATTTATCTCTTAAACTGTGCTTGTGCAACTGTGATTAACG 480
608 TCCAGTCCCGATCGTTATTTATCTCTTAAACTGTGCTTGTGCAACTGTGATTAACG 667
481 AACGTGATCGACGCGATCTCTACAGAGGAGAGGATGCAATGCTGATCTTACCGGCGAT 540
668 AACGTGATCGACGCGATCTCTACAGAGGAGAGGATGCAATGCTGATCTTACCGGCGAT 727
541 CCGCAAAAGCGCGTTTGGCGCACTGGAACGGGTGCTTAATTTTCCGCAATCAACTGTGC 600
728 CCGCAAAAGCGCGTTTGGCGCACTGGAACGGGTGCTTAATTTTCCGCAATCAACTGTGC 787
601 CTTAAACGTGAGAAACAGACGAAAGCTGTTCAATTAACGAGCATTAACCATGGAATC 660
788 CTTAAACGTGAGAAACAGACGAAAGCTGTTCAATTAACGAGCATTAACCATGGAATC 847
661 AAGGTGAGCGCGCAATGTCATTAACCGGTGCGTAAAGCTCGCATCAATGCTGAGC 720
848 AAGGTGAGCGCGCAATGTCATTAACCGGTGCGTAAAGCTCGCATCAATGCTGAGC 907
721 GAGATATTTCTCTGCAACAGCAGAGGAATGCCGAGCGCGGTGTGGGAAGGATACG 780
908 GAGATATTTCTCTGCAACAGCAGAGGAATGCCGAGCGCGGTGTGGGAAGGATACG 967
781 GATTACACAGAGGAAACACCTTGCTAAGTTTGATTAACGCGCAATTTTATTGCTACAA 840
968 GATTACACAGAGGAAACACCTTGCTAAGTTTGATTAACGCGCAATTTTATTGCTACAA 1027
841 CGCACGCGCAGAGGTGGCGCGAGCGCGCAACCCCGTTATTAAGTTTATCAAGACGCG 900
1028 CGCACGCGCAGAGGTGGCGCGAGCGCGCAACCCCGTTATTAAGTTTATCAAGACGCG 1087
901 TTGACGCGCCCATTCACCGCAAAACAGCGGTATGTGATTAACCATTCAGTGTCTG 960
1088 TTGACGCGCCCATTCACCGCAAAACAGCGGTATGTGATTAACCATTCAGTGTCTG 1147
961 TTATATGCGCGGACAGATTAATCTGCAAAATCTGCGCGCGCGCATGAGACTCAACTG 1020
1148 TTATATGCGCGGACAGATTAATCTGCAAAATCTGCGCGCGCGCATGAGACTCAACTG 1207

1021 ACGTTCCCGGTCAGCGGATTAACAGCGCGCAGGTGTGAACTGTGTTGAACGCTGG 1080
1208 ACGTTCCCGGTCAGCGGATTAACAGCGCGCAGGTGTGAACTGTGTTGAACGCTGG 1267
1081 CGTGGCTTAAGGATTAACAGCGGATTAACAGCGGATTAACAGCGGATTAACAGCG 1140
1268 CGTGGCTTAAGGATTAACAGCGGATTAACAGCGGATTAACAGCGGATTAACAGCG 1327
1141 CAGATGCGGTATTAACAGCGGATTAACAGCGGATTAACAGCGGATTAACAGCG 1200
1328 CAGATGCGGTATTAACAGCGGATTAACAGCGGATTAACAGCGGATTAACAGCG 1387
1201 CTGGCAGATGTGAAGAGGAAATGCGCAGGATGTGTTGCTGGCAGATTTACGCA 1260
1388 CTGGCAGATGTGAAGAGGAAATGCGCAGGATGTGTTGCTGGCAGATTTACGCA 1447
1261 ATCGTGAATGAAGACGATTAACCGCTTGCAGTTGTAA 1299
1448 ATCGTGAATGAAGACGATTAACCGCTTGCAGTTGTAA 1486

RESULT 3
US-10-156-660-3
; Sequence 3, Application US/10156660
; Publication No. US20030103958A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay M.
; APPLICANT: Kretz, Keith
; APPLICANT: Gray, Kevin A.
; APPLICANT: Barton, Nelson R.
; APPLICANT: Garrett, James B.
; APPLICANT: O'Donoghue, Eileen
; TITLE OF INVENTION: PHYLASES, NUCLEIC ACIDS ENCODING THEM
; FILE REFERENCE: 09010-029007
; CURRENT APPLICATION NUMBER: US/10/156,660
; CURRENT FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: US 09/866,379
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: US 09/580,515
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 09/318,528
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: US 09/291,931
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: US 09/259,214
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: US 08/910,798
; PRIOR FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1901
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (188) ... (1483)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 403
; OTHER INFORMATION: n = A,T,C or G
US-10-156-660-3

Query Match 99.6%; Score 1293.2; DB 15; Length 1901;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1295; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

1 ATGAAAGCCATCTTAATCCATTTTATCTCTTGTATCCGTAAACCCCGAATCTGCA 60
188 ATGAAAGCCATCTTAATCCATTTTATCTCTTGTATCCGTAAACCCCGAATCTGCA 247

QY 61 TTCCGCTCAGAGTGAAGCCGGAAGCTGAAGCTGGAAGAGTGGTGTGATGTCAGTGTGATGCT 120
DB 248 TTCCGCTCAGAGTGAAGCCGGAAGCTGGAAGAGTGGTGTGATGTCAGTGTGATGCT 307
QY 121 GTGGCGTGTCCAAACCAAGGCGCAAGCAATGAGATGTCAGATGTCAGATGTCAGATGTCAG 180
DB 308 GTGGCGTGTCCAAACCAAGGCGCAAGCAATGAGATGTCAGATGTCAGATGTCAGATGTCAG 367
QY 181 ACCGCGCGGTAAACCTGGGTGGCTGA CACCGCGCGGTGGTGGTGAATGCTAATGCTATGTC 240
DB 368 ACCGCGCGGTAAACCTGGGTGGCTGA CACCGCGCGGTGGTGGTGAATGCTAATGCTATGTC 427
QY 241 GGAATTAACCAACGCAAGCTGTGGTACCGAATGCTGGCGAATAAGGCGTGGCCG 300
DB 428 GGAATTAACCAACGCAAGCTGTGGTACCGAATGCTGGCGAATAAGGCGTGGCCG 487
QY 301 CAGTGTGGTCAAGGTGGCGATTAATGCTGAATGTCAGAGCGTAACCGTAAACAGCGCA 360
DB 488 CAGTGTGGTCAAGGTGGCGATTAATGCTGAATGTCAGAGCGTAACCGTAAACAGCGCA 547
QY 361 GCGTTTCGCGCGCGGCTGGCACTGATGTCAGATTAACCGTAACATTAACCGCAATACG 420
DB 548 GCGTTTCGCGCGCGGCTGGCACTGATGTCAGATTAACCGTAACATTAACCGCAATACG 607
QY 421 TCCAGTCCCGAATCCGTTATTTAATCCTCTAATAACCTGGCGTTGGCACTGGATTAACGG 480
DB 608 TCCAGTCCCGAATCCGTTATTTAATCCTCTAATAACCTGGCGTTGGCACTGGATTAACGG 667
QY 481 AACGTGACTGACGCGATCCTCAGACGAGGAGGAGGATGCTGACTTTACCGGGCAT 540
DB 668 AACGTGACTGACGCGATCCTCAGACGAGGAGGAGGATGCTGACTTTACCGGGCAT 727
QY 541 CGGCAAAACGCGGTTTCGCGAACTGAAACGCGGTCTTAATTTTCGCAATCAACTTGTGC 600
DB 728 CGGCAAAACGCGGTTTCGCGAACTGAAACGCGGTCTTAATTTTCGCAATCAACTTGTGC 787
QY 601 CTTAAACGTAAGAAACAGGACGAAAGCTGTATTAAGCAGGCAATTAACATCGGAATC 660
DB 788 CTTAAACGTAAGAAACAGGACGAAAGCTGTATTAAGCAGGCAATTAACATCGGAATC 847
QY 661 AAGGTGAGCGCGCAATGCTCATTTAACCGGTGGGTAAAGCTGCAATCAATGCTGACG 720
DB 848 AAGGTGAGCGCGCAATGCTCATTTAACCGGTGGGTAAAGCTGCAATCAATGCTGACG 907
QY 721 GAGATATTTCTCTGCAACAGCAAGCAAGGAAATGCGGAGCGGGGTGGGAAAGATCAC 780
DB 908 GAGATATTTCTCTGCAACAGCAAGCAAGGAAATGCGGAGCGGGGTGGGAAAGATCAC 967
QY 781 GATTCAACACAGTGGAAACCTTGCTAATGTTGCAATTAAGCGCAATTTATTTGCTACAA 840
DB 968 GATTCAACACAGTGGAAACCTTGCTAATGTTGCAATTAAGCGCAATTTATTTGCTACAA 1027
QY 841 CGCAGCGCAAGAGTTGCGCGAGCGCGCACCGCCGTTATTAATTTGATCAAGACAGG 900
DB 1028 CGCAGCGCAAGAGTTGCGCGAGCGCGCACCGCCGTTATTAATTTGATCAAGACAGG 1087
QY 901 TTGAGCGCCCATTCACCGCAAAAACAGGCGTATGTTGATCAATTAACCTTCAAGTCTG 960
DB 1088 TTGAGCGCCCATTCACCGCAAAAACAGGCGTATGTTGATCAATTAACCTTCAAGTCTG 1147
QY 961 TTTATGCGCGGACAGATTAATCTAATCTGGCAATCTGGGCGGCACTGAGGCTCAACTG 1020
DB 1148 TTTATGCGCGGACAGATTAATCTAATCTGGCAATCTGGGCGGCACTGAGGCTCAACTG 1207
QY 1021 ACGCTTCCCGGTCAAGCGGATTAACAGCGCGCAAGTGTGAACCTGTTGTTGAACGCTGG 1080
DB 1208 ACGCTTCCCGGTCAAGCGGATTAACAGCGCGCAAGTGTGAACCTGTTGTTGAACGCTGG 1267
QY 1081 CGTGGCTTAAGGATTAACAGCGGATTAACAGGTTTGGTGTCTTCAAGCTTTACG 1140
DB 1268 CGTGGCTTAAGGATTAACAGCGGATTAACAGGTTTGGTGTCTTCAAGCTTTACG 1327
QY 1141 CAGATCGGTATTAACGCGGCTGTGATTAATTAATAGCGCGCGGAGAGGTGAACGAC 1200

DB 1228 CAGATCGGTATTAACGCGGCTGTGATTAATACCGCGCGGAGAGGTGAACGAC 1387
QY 1201 CTGGCAGGATGTAAGAGCGAAATGCGAGGCGATGTTGTTGCGAGGTTTACGCA 1280
DB 1388 CTGGCAGGATGTAAGAGCGAAATGCGAGGCGATGTTGTTGCGAGGTTTACGCA 1447
QY 1281 ATCGTAATGAAGCAGCATACCGCTTGGCATTTGTA 1299
DB 1448 ATCGTAATGAAGCAGCATACCGCGCTGCAATTGTA 1486

RESULT 4
US-10-601-319-7
; Sequence 7, Application US/10601319
; Publication No. US20040091968A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay M.
; APPLICANT: Kretz, Keith A.
; APPLICANT: Gray, Kevin A.
; APPLICANT: Barton, Nelson Robert
; APPLICANT: Garrett, James B.
; APPLICANT: O'Donoghue, Eileen
; APPLICANT: Machur, Eric J.
; TITLE OF INVENTION: RECOMBINANT PHYTASES AND METHODS OF MAKING
; TITLE OF INVENTION: AND USING THEM
; FILE REFERENCE: 09010-029011
; CURRENT APPLICATION NUMBER: US/10/601,319
; PRIOR FILING DATE: 2003-06-20
; PRIOR APPLICATION NUMBER: US 09/866,379
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: US 09/580,515
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 09/318,528
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: US 09/291,931
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: US 09/259,214
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: US 08/910,798
; PRIOR FILING DATE: 1997-08-13
; PRIOR APPLICATION NUMBER: US 08/910,798
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1901
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (188)...(1483)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 403
; OTHER INFORMATION: n = A,T,C or G
; US-10-601-319-7

Query Match 99.6%; Score 1293.2; DB 17; Length 1901;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1995; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATGAAGCAGATCTTAATCCATTTTATCTCTTGAATTCGGTTAAACCCGCAATCTGA 60
DB 188 ATGAAGCAGATCTTAATCCATTTTATCTCTTGAATTCGGTTAAACCCGCAATCTGA 247
QY 61 TTCCGCTCAGAGTGAAGCCGGAAGCTGGAAGAGTGGTGTGATGTCAGTGTGATGCT 120
DB 248 TTCCGCTCAGAGTGAAGCCGGAAGCTGGAAGAGTGGTGTGATGTCAGTGTGATGCT 307
QY 121 GTGGCGTGTCCAAACCAAGGCGCAAGCAATGAGATGTCAGATGTCAGATGTCAGATGTCAG 180
DB 308 GTGGCGTGTCCAAACCAAGGCGCAAGCAATGAGATGTCAGATGTCAGATGTCAGATGTCAG 367
QY 181 ACCGCGCGGTAAACCTGGGTGGCTGA CACCGCGCGGTGGTGGTGAATGCTAATGCTATGTC 240

DB 368 ACCGCGCCGCTAAGACGAGGCTGACACCGCGAGGCTGAGCTTAATCCCTTATCTC 427
DB 241 GGACATTTACCAACGCCAGCGCTGTGTAGCGACGAGATTGCTGCGAAGAAAGGCTGCCG 300
DB 428 GGAATTTACCAACGCCAGCGCTGTGTAGCGACGAGATTGCTGCGAAGAAAGGCTGCCG 487
DB 301 CAGCTGTGTAGGTCGCGATTTATTTGCTGATGTGACGAGCGTACCCGTTAAACAGCGGAA 360
DB 488 CAGCTGTGTAGGTCGCGATTTATTTGCTGATGTGACGAGCGTACCCGTTAAACAGCGGAA 547
DB 361 GCGTTGCGCGCGCGCGCTGCGACCTGCACTGCAATTAACGCTACATCCAGGAGATAG 420
DB 548 GCGTTGCGCGCGCGCGCTGCGACCTGCACTGCAATTAACGCTACATCCAGGAGATAG 607
DB 421 TCCAGTCCCGATCCGTTAATTTAATCTTAAACTGCGCTTTGCAATGATTAACGCG 480
DB 608 TCCAGTCCCGATCCGTTAATTTAATCTTAAACTGCGCTTTGCAATGATTAACGCG 667
DB 481 AACGTGATGACCGGATCTCTGACGAGGCGAGAGGCTCAATTGCTGACTTTACCGGCGAT 540
DB 668 AACGTGATGACCGGATCTCTGACGAGGCGAGAGGCTCAATTGCTGACTTTACCGGCGAT 727
DB 541 CCGCAAAACGCGCTTTGCGCGAATCTGGAACGCGGCTTTAATTTCCGCAATCAAACTTGTG 600
DB 728 CCGCAAAACGCGCTTTGCGCGAATCTGGAACGCGGCTTTAATTTCCGCAATCAAACTTGTG 787
DB 601 CTTAAACGTGAGAAACAGACGAAAGCTGTTCAATTAACGAGGATTAACGATCGAATC 660
DB 788 CTTAAACGTGAGAAACAGACGAAAGCTGTTCAATTAACGAGGATTAACGATCGAATC 847
DB 661 AAGGTGAGCGCGCAATGTTCTAATTAACGCTGCGGTAAGCTTGCATATGCTGACG 720
DB 848 AAGGTGAGCGCGCAATGTTCTAATTAACGCTGCGGTAAGCTTGCATATGCTGACG 907
DB 721 GAGATATTTCTCTGCAACAGACGAGGAAATGCGGAGCGCGGCTGCGGAGAGATACC 780
DB 908 GAGATATTTCTCTGCAACAGACGAGGAAATGCGGAGCGCGGCTGCGGAGAGATACC 967
DB 781 GATTACACAGGAGGAAACCTTGTCTAAGTTTGATTAACGCAATTTAATTTGCTCAAA 840
DB 968 GATTACACAGGAGGAAACCTTGTCTAAGTTTGATTAACGCAATTTAATTTGCTCAAA 1027
DB 841 CGCACGCGCAGAGGTTGCGCGACGCGCGCAACCCGTTAATTAAGTTTGAATCAGACGCG 900
DB 1028 CGCACGCGCAGAGGTTGCGCGACGCGCGCAACCCGTTAATTAAGTTTGAATCAGACGCG 1087
DB 901 TTGACGCGCCCATCAACCGCAAAACAGCGGATGATGATTAACCACTTCAGTGTG 960
DB 1088 TTGACGCGCCCATCAACCGCAAAACAGCGGATGATGATTAACCACTTCAGTGTG 1147
DB 961 TTATATGCGCGGACAGATTAATCTGCAAAATCTCGGCGCGGCACTGAGACTCACTG 1020
DB 1148 TTATATGCGCGGACAGATTAATCTGCAAAATCTCGGCGCGGCACTGAGACTCACTG 1207
DB 1021 ACCTGTCGCGGTCAACCGGATTAACGCGCGCAGAGTGTGAATCTGTGTTGAACGCTG 1080
DB 1208 ACCTGTCGCGGTCAACCGGATTAACGCGCGCAGAGTGTGAATCTGTGTTGAACGCTG 1267
DB 1081 CGTGTGCTAAGCGATTAACGCGCAATGATTAAGTTTGTGCTGCTTCAACCTTTACG 1140
DB 1268 CGTGTGCTAAGCGATTAACGCGCAATGATTAAGTTTGTGCTGCTTCAACCTTTACG 1327
DB 1141 CAGATGTGCTAATTAACGCGCGCTGTCAATTAATACGCGCGCGAGAGAGTGAATCAGC 1200
DB 1328 CAGATGTGCTAATTAACGCGCGCTGTCAATTAATACGCGCGCGAGAGAGTGAATCAGC 1387
DB 1201 CTGCGCAGAGTGTGAAGAGCGAAATGCGCAGGCGCAATGTGTTGTTGACGAGTTTACGCA 1260
DB 1388 CTGCGCAGAGTGTGAAGAGCGAAATGCGCAGGCGCAATGTGTTGTTGACGAGTTTACGCA 1447
DB 1261 ATCGTGAATGAAGCAGATACCGGCTTGCAGTTTGTAA 1299

DB 1448 ATCGTGAATGAAGCAGATACCGGCGTGCAGTTTGTAA 1486
RESULT 5
US-09-866-379-5
; Sequence 5, Application US/09866379
; Patent No. US20020136754A1
; GENERAL INFORMATION:
; APPLICANT: DIVERSA CORPORATION
; APPLICANT: SHORT, Jay
; APPLICANT: KRETZ, Keith
; APPLICANT: GRAY, Kevin
; APPLICANT: BARTON, Nelson
; APPLICANT: GARRETT, James
; APPLICANT: O'DONOGHUE, Eileen
; TITLE OF INVENTION: RECOMBINANT BACTERIAL PHYTASES AND USES THEREOF
; FILE REFERENCE: DIVER1370-7
; CURRENT APPLICATION NUMBER: US/09/866,379
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: US 09/580,515
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 09/318,528
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: US 09/291,931
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: US 09/259,214
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: US 08/910,798
; PRIOR FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 1901
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(1901)
; OTHER INFORMATION: n is any nucleotide
US-09-866-379-5
Query Match 99.4%; Score 1291.6; DB 9; Length 1901;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1294; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
DB 1 ATGAAGCCATCTTAATCCCAATTTTATCTTTCTGATTCGCTTAACCCGCAATCTGCA 60
DB 188 ATGAAGCCATCTTAATCCCAATTTTATCTTTCTGATTCGCTTAACCCGCAATCTGCA 247
DB 61 TTGCTCAAGTGAAGCCGCGAGCTGAAGCTGGAAGTGTGATGTCAGTCCGATAGT 120
DB 248 TTGCTCAAGTGAAGCCGCGAGCTGAAGCTGGAAGTGTGATGTCAGTCCGATAGT 307
DB 121 GTGCGTCTCAACAGGCGCAAGCACTGATGCAAGGATGCAACCCAGAGCGATGCGCA 180
DB 308 GTGCGTCTCAACAGGCGCAAGCACTGATGCAAGGATGCAACCCAGAGCGATGCGCA 367
DB 181 ACCTGCGCGGTAAACTGAGTTGCTGACACCGCGCGTGTGATGCTAATGCTTATCTC 240
DB 368 ACCTGCGCGGTAAACTGAGTTGCTGACACCGCGCGTGTGATGCTAATGCTTATCTC 427
DB 241 GGAATTTACCAACGCCAGCGCTGTGTAGCGACGAGATTGCTGCGAAGAAAGGCTGCCG 300
DB 428 GGAATTTACCAACGCCAGCGCTGTGTAGCGACGAGATTGCTGCGAAGAAAGGCTGCCG 487
DB 301 CAGTCTGTGATGAGTGTGATTAATGCTGATGTGACGAGGATACCCGTTAAACAGCGGAA 360
DB 488 CAGTCTGTGATGAGTGTGATTAATGCTGATGTGACGAGGATACCCGTTAAACAGCGGAA 547
DB 361 GCGTTGCGCGCGCGGCTGCGACCTGACTGTGCAATTAACGTAATCAACCGAGGATAGC 420
DB 548 GCGTTGCGCGCGCGGCTGCGACCTGACTGTGCAATTAACGTAATCAACCGAGGATAGC 607

QY	421	CCGATCCCGGATCCGTTATTTTAAATCCCTCTAAAAAATGGCGTTTGGCACTGGATAACGGG	480
Db	608	TCGATCCCGGATCCGTTATTTATATCTCTAAAAAATGGCGTTTGGCACTGGATAACGGG	667
QY	481	AACGTGACTGACGCGATCTCTACGACGGGCAAGGAGGTCAATTGTGACTTTACCGGCGCAT	540
Db	668	AACGTGACTGACGCGATCTCTACGACGGGCAAGGAGGTCAATTGTGACTTTACCGGCGCAT	727
QY	541	CGGCAAAAGGGGTTTTCGGGAATGGAAAGGGGGCTTAATTTTCGCGCATCAAACTTTGGC	600
Db	728	CGGCAAAAGGGGTTTTCGGGAATGGAAAGGGGGCTTAATTTTCGCGCATCAAACTTTGGC	787
QY	601	CTTAAACCTGAGAAAACAGGACGAAAGCTGTTCTATTAAACGACAGCATTTACATCGGAACTC	660
Db	788	CTTAAACCTGAGAAAACAGGACGAAAGCTGTTCTATTAAACGACAGCATTTACATCGGAACTC	847
QY	661	AAGGTGAGGGCCGCAATATGTTCTCATTTAACCGGTGGCGGTAAAGCTTGGCATTAATGCTGACG	720
Db	848	AAGGTGAGGGCCGCAATATGTTCTCATTTAACCGGTGGCGGTAAAGCTTGGCATTAATGCTGACG	907
QY	721	GAGATATTTCTCTCTCAACAAAGCAACAGGGAATGCCGAGACCGGGGTGGGGAAGATATCAC	780
Db	908	GAGATATTTCTCTCTCAACAAAGCAACAGGGAATGCCGAGACCGGGGTGGGGAAGATATCAC	967
QY	781	GATTTCACAACCACTGGAAACACTTGTGTAAGTTTGCAATACCGGCAATTTTATTTGCTACAA	840
Db	968	GATTTCACAACCACTGGAAACACTTGTGTAAGTTTGCAATACCGGCAATTTTATTTGCTACAA	1027
QY	841	CGCAAGCCAGAGGGTTGCCCGGACGGCGGCAACCCGGTTATTTGATTTGATCAAGACAGCG	900
Db	1028	CGCAAGCCAGAGGGTTGCCCGGACGGCGGCAACCCGGTTATTTGATTTGATCAAGACAGCG	1087
QY	901	TTGACGCCCCATCCAACCGCAAAAACAAGCGTATGTTGACATTAACCACTTCAAGTGTG	960
Db	1088	TTGACGCCCCATCCAACCGCAAAAACAAGCGTATGTTGACATTAACCACTTCAAGTGTG	1144
QY	961	TTTATTCGCCCGGACAAGATCTAATCTGGGAATTCGCGGGGCACTGGAGGCTCAACTGG	1021
Db	1148	TTTATTCGCCCGGACAAGATCTAATCTGGGAATTCGCGGGGCACTGGAGGCTCAACTGG	1207
QY	1021	ACGCTTCCCGGTCAGCCGGATTAACGCCGCCAGGTGTGAACTGGTGTTTTGAACGCTGG	1088
Db	1208	ACGCTTCCCGGTCAGCCGGATTAACGCCGCCAGGTGTGAACTGGTGTTTTGAACGCTGG	1266
QY	1081	CGTCCGCTAAGCGATTAACGCCCACTGTGATTCAGGTTTCGCTGTCCTTCCAGACTTTACAG	1144
Db	1268	CGTCCGCTAAGCGATTAACGCCCACTGTGATTCAGGTTTCGCTGTCCTTCCAGACTTTACAG	1322
QY	1141	CAGATGCGGTAAAAAACGCCGCTGTCAATTAAATACGCCGCCCGGAGAGGTTAAACTGACC	1200
Db	1328	CAGATGCGGTAAAAAACGCCGCTGTCAATTAAATACGCCGCCCGGAGAGGTTAAACTGACC	1387
QY	1201	CTGGCAGAGTGTGAAGAGGAAATGCGCAGGGCATATGTTCGTTGGCAGGTTTACGCAA	1266
Db	1388	CTGGCAGAGTGTGAAGAGGAAATGCGCAGGGCATATGTTCGTTGGCAGGTTTACGCAA	1444
QY	1261	ATCGTAATGAAGCAACGATACCCCGTTCAGTTTGTAA	1299
Db	1448	ATCGTAATGAAGCAACGATACCCCGTTCAGTTTGTAA	1486

RESULT 6
US-09-866-379-9
Sequence 9, Application US/09866379
Patent No. US20020136754A1
GENERAL INFORMATION:
APPLICANT: DIVERSA CORPORATION
APPLICANT: SHORT, Jay
APPLICANT: KRETTZ, Keith
APPLICANT: GRAY, Kevin
APPLICANT: BARTON, Nelson
APPLICANT: GARRETT, James
APPLICANT: O'DONOGHUE, Eileen

```

1 TITLE OF INVENTION: RECOMBINANT BACTERIAL PHATASES AND USES THEREOF
2
3 FILE REFERENCE: DIVER1370-7
4
5 CURRENT APPLICATION NUMBER: US/09/866,379
6
7 CURRENT FILING DATE: 2001-05-24
8
9 PRIOR APPLICATION NUMBER: US 09/580,515
10
11 PRIOR FILING DATE: 2000-05-25
12
13 PRIOR APPLICATION NUMBER: US 09/318,528
14
15 PRIOR FILING DATE: 1999-05-25
16
17 PRIOR APPLICATION NUMBER: US 09/291,931
18
19 PRIOR FILING DATE: 1999-04-13
20
21 PRIOR APPLICATION NUMBER: US 09/259,214
22
23 PRIOR FILING DATE: 1999-03-01
24
25 PRIOR APPLICATION NUMBER: US 08/910,798
26
27 PRIOR FILING DATE: 1997-08-13
28
29 NUMBER OF SEQ ID NOS: 10
30
31 SOFTWARE: Patentin version 3.1
32
33 SEQ ID NO 9
34
35 LENGTH: 1901
36
37 TYPE: DNA
38
39 ORGANISM: Escherichia coli
40
41 FEATURE:
42
43 NAME/KEY: misc_feature
44
45 LOCATION: (1)..(1901)
46
47 OTHER INFORMATION: n is any nucleotide
48
49 US-09-866-379-9

```

	Query Match	99.4%;	Score 1291.6;	DB 9;	Length 1901;	
	Best Local Similarity	99.6%;	Pred. No. 0;			
	Matches 1294;	Conservative	0;	Mismatches	5;	Indels
						Gaps 0;
OY	1 ATGAAAGCCATCTTAAATGCCATTTTATACTTCCTTGATTCCGGTTAACCCCGCAATCTGCA	60				
Db	188 ATGAAGAAGCATCTTAATCCTAATTCCTTTATCTCTTGTGATTCGGTTAACCCGCMAATCTCA	247				
OY	61 TTGCCTCAGATGACCGCGAGCTGAAGGTGGAAGTAGTGATGTGACTGCTGATGAT	120				
Db	248 TTCCTCAGATGACCGCGAGCTGAAGTAGTGATGTGATGTGATGTGATGTGATGTGATGTGAT	307				
OY	121 GTGCGTGTCCAAACCAAAGGCCACAATGACATGCAAGATGTCAACCCAGACGATGCGCA	180				
Db	308 GTGCGTGTCCAAACCAAAGGCCACAATGACATGCAAGATGTGCAACCCAGACGATGCGCA	367				
OY	181 ACCGAGCCCGGTAAAACTGGGTGGCTGACAACGCGCGGTGGTGAAGCTAATCGACTATCTC	240				
Db	368 AACCAGCCCGGTAAAACTGGGTGGCTGACAACGCGCGGTGGTGAAGCTAATCGACTATCTC	427				
OY	241 GGACATTAACCAACGCGCACGCGTCTGGTAGCCGACGCGATTGCTGGCAAAAAGGGCTGCCG	300				
Db	428 GGAATTAACCAACGCGCACGCGTCTGGTAGCCGACGCGATTGCTGGCAAAAAGGGCTGCCG	487				
OY	301 CAGTCTGGTCAAGTTCGCGATTAATTGCTGATGTCGACGAGCGGTACCCGTAAAAACAGCGCA	360				
Db	488 CAGTCTGGTCAAGTTCGCGATTAATTGCTGATGTCGACGAGCGGTACCCGTAAAAACAGCGCA	547				
OY	361 GCCTTCGCGCGCGGGCTGGCACTGACTGTGCATTAACCGTACATTAACCGAGCAGATAAG	420				
Db	548 GCCTTTCGCGCGCGGGCTGGCACTGACTGTGCATTAACCGTACATTAACCGAGCAGATAAG	607				
OY	421 TCCAGTCCCCATCCGTTATTTAATTCCTTTAAAACTGGCGTTTTGGCACTGGATAAGCGS	480				
Db	608 TCAGATCCCAGTCCGTTATTTAATTCCTTTAAAACTGGCGTTTTGGCACTGGATAAGCGS	667				
OY	481 AACGTGACTGACGCGATCTCAGACGAGCAGAGAGGTCAATTGCTGACTTTTACCGGGCAT	540				
Db	668 AACGTGACTGACGCGATCTCAGACGAGCAGAGAGGTCAATTGCTGACTTTTACCGGGCAT	727				
OY	541 CGGCAAAAGCGCGTTTCGCGAACTGGAACGGGTGCTTAATTTTCCGCAATCAAACTTGTGC	600				
Db	728 CGGCAAAAGCGCGTTTCGCGAACTGGAACGGGTGCTTAATTTTCCGCAATCAAACTTGTGC	787				
OY	601 CTAAACCTGAGAAACAGAACGAAAGCTGTTCATTAACGACAGGCAATTAACATCGGAATCTC	660				
Db	788 CTTAAACCTGAGAAACAGAACGAAAGCTGTTCATTAACGACAGGCAATTAACATCGGAATCTC	847				

Qy	661	AAGGTGAGCGCCGACAATGTCATTTAACCGGTCGGGTAACCTTCGATCAATGCTACG	720
Db	848	AAGGTGAGCGCCGACAATGTCATTTAACCGGTCGGGTAACCTTCGATCAATGCTACG	907
Qy	721	GAGATATTTCTCCTGCAACAAGCA CAGGGAAATGCCGAGCCGGGTGGGAGGATCAC	780
Db	908	GAGATATTTCTCCTGCAACAAGCA CAGGGAAATGCCGAGCCGGGTGGGAGGATCAC	967
Qy	781	GATTTCACACCGATGGAACACTTGTGTAAGTTTGCAATACGGGCAATTTATTTGCTACAA	840
Db	968	GATTTCACACCGATGGAACACTTGTGTAAGTTTGCAATACGGGCAATTTATTTGCTACAA	1027
Qy	841	CGCAGCGCAGAGGTTGGCCCGCAGCCGCGCA CCCCCTGATTAAGATTTGATCAAGACAGCG	900
Db	1028	CGCAGCGCAGAGGTTGGCCCGCAGCCGCGCA CCCCCTGATTAAGATTTGATCAAGACAGCG	1087
Qy	901	TTGAACGCCCATTCACACCGCAAAAACAGGCGATAGGTGTGACATTAACCACTTCAGTGTG	960
Db	1088	TTGAACGCCCATTCACACCGCAAAAACAGGCGATAGGTGTGACATTAACCACTTCAGTGTG	1147
Qy	961	TTTATATCCCGGCAACAGATATCTATCTGGCAATCTCCGCGCGCGCATGAGACTCAACTG	1020
Db	1148	TTTATATCCCGGCAACAGATATCTATCTGGCAATCTCCGCGCGCGCATGAGACTCAACTG	1207
Qy	1021	ACGCTTCCCGGCTCAGCCGGATTAACACGCGCCAGCGTGTGAACTGTGTTGAAACGCTG	1080
Db	1208	ACGCTTCCCGGCTCAGCCGGATTAACACGCGCGCCAGCGTGTGAACTGTGTTGAAACGCTG	1267
Qy	1081	CGTGGCTTAAGCGATTAACAGCCAGTGAATTCAGTTTGGCTGTCTTCCACATTTACAG	1140
Db	1268	CGTGGCTTAAGCGATTAACAGCCAGTGAATTCAGTTTGGCTGTCTTCCACATTTACAG	1327
Qy	1141	CAGATGCTGTATTAACACCCCGCTGCATTTAAATACCGCCCGGAGAGGTAAACTGAC	1200
Db	1328	CAGATGCTGTATTAACACCCCGCTGCATTTAAATACCGCCCGGAGAGGTAAACTGAC	1387
Qy	1201	CTGGCAGAGTGTGAAGACGCAAAATCGCAGGGCATGTTCGTTGGCAGGTTTACGCA	1260
Db	1388	CTGGCAGAGTGTGAAGACGCAAAATCGCAGGGCATGTTCGTTGGCAGGTTTACGCA	1447
Qy	1261	ATCGTGAATGAAGCAGCATACCCGCTTGACGTTTGAA	1299
Db	1448	ATCGTGAATGAAGCAGCATACCCGCTTGACGTTTGAA	1486

QY 841 CGCAGCCGAGAGTTGCCCCGACGCCGCCGATTTAGATTGATGAAGACACCG 900
DB 1028 CGCAGCCGAGAGTTGCCCCGACGCCGCCGATTTAGATTGATGAAGACACCG 1087
QY 901 TTGAGCCGCCATCCACCGCAAAAAGGCGATGCTGATGACATTCACCACTTCAGTGTG 960
DB 1088 TTGAGCCGCCATCCACCGCAAAAAGGCGATGCTGATGACATTCACCACTTCAGTGTG 1147
QY 961 TTTATCGCCGACACGATATCTAATCTGCAAAATCTCGCGCGCGCATCTGAGCTCACTGG 1020
DB 1148 TTTATCGCCGACACGATATCTAATCTGCAAAATCTCGCGCGCGCATCTGAGCTCACTGG 1207
QY 1021 AGCTTCCCGGTACACCGGATTAACACGCCGCCGAGGTGTAATCTGCTTTTAAAGCTGG 1080
DB 1208 AGCTTCCCGGTACACCGGATTAACACGCCGCCGAGGTGTAATCTGCTTTTAAAGCTGG 1267
QY 1081 CGTGGCTTAAGGATTAACAGCCAGTGAATTCAGTTCGCTGATCTTCAGACTTTACAG 1140
DB 1268 CGTGGCTTAAGGATTAACAGCCAGTGAATTCAGTTCGCTGATCTTCAGACTTTACAG 1327
QY 1141 CAGATGCGTGAATAACCGCGCTGCTAATTAATAGCGCGCGCGAGAGGTGAACCTGACC 1200
DB 1328 CAGATGCGTGAATAACCGCGCTGCTAATTAATAGCGCGCGCGAGAGGTGAACCTGACC 1387
QY 1201 CTGGCAGGATGTAAGAGGGAATGCGCAGGCGATGTTGTTGGCAGGTTTACGCA 1260
DB 1388 CTGGCAGGATGTAAGAGGGAATGCGCAGGCGATGTTGTTGGCAGGTTTACGCA 1447
QY 1261 ATCGTAATGAAGACGACATACCGCGCTTGCAGTTGTAA 1299
DB 1448 ATCGTAATGAAGACGACATACCGCGCTTGCAGTTGTAA 1486

RESULT 8

US-09-866-379-6

Sequence 6, Application US/09866379

Patent No. US20020136754A1

GENERAL INFORMATION:

APPLICANT: DIVERSA CORPORATION

APPLICANT: SHORT, Jay

APPLICANT: KRETZ, Keith

APPLICANT: GRAY, Kevin

APPLICANT: BARTON, Nelson

APPLICANT: GARRETT, James

APPLICANT: O'DONOGHUE, Eileen

TITLE OF INVENTION: RECOMBINANT BACTERIAL PHYTASES AND USES THEREOF

FILE REFERENCE: DIVER1370-7

CURRENT APPLICATION NUMBER: US/09/866,379

CURRENT FILING DATE: 2001-05-24

PRIOR APPLICATION NUMBER: US 09/580,515

PRIOR FILING DATE: 2000-05-25

PRIOR APPLICATION NUMBER: US 09/318,528

PRIOR FILING DATE: 1999-05-25

PRIOR APPLICATION NUMBER: US 09/291,931

PRIOR FILING DATE: 1999-04-13

PRIOR APPLICATION NUMBER: US 09/259,214

PRIOR FILING DATE: 1999-03-01

PRIOR APPLICATION NUMBER: US 08/910,798

PRIOR FILING DATE: 1997-08-13

NUMBER OF SEQ ID NOS: 10

SOFTWARE: PatentIn version 3.1

SEQ ID NO 6

LENGTH: 1901

TYPE: DNA

ORGANISM: Escherichia coli

FEATURE:

NAME/KEY: misc_feature

LOCATION: (1)..(1901)

OTHER INFORMATION: n is any nucleotide

US-09-866-379-6

Query Match 99.2%; Score 1288.4; DB 9; Length 1901;

Best Local Similarity 99.5%; Pred. No. 0;
Matches 1292; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ATGAAAGCATCTTAATCCCATTTTATCTTTGATTCGGTAACCCCGCAATCTGCA 60
DB 188 ATGAAAGCATCTTAATCCCATTTTATCTTTGATTCGGTAACCCCGCAATCTGCA 247
QY 61 TTGCGTCAGATGAGCCGGAAGCTGAAGTGAAGTGTGATTCAGTCTCATGT 120
DB 248 TTGCGTCAGATGAGCCGGAAGCTGAAGTGTGATTCAGTCTCATGT 307
QY 121 GTGCGTCTCCCAACCAAGGCCAGCACTGATGAGATGTACCCAGACGATGCGCA 180
DB 308 GTGCGTCTCCCAACCAAGGCCAGCACTGATGAGATGTACCCAGACGATGCGCA 367
QY 181 ACTGCGCGGTAAACTGGGTGCTGACACCGCGCGGTGTTGAGCTAATGCGCTATCTC 240
DB 368 ACTGCGCGGTAAACTGGGTGCTGACACCGCGCGGTGTTGAGCTAATGCGCTATCTC 427
QY 241 GGAATTAACCAAGCCAGCGCTGCTGATGCGAGATGCTGCGCAAAAAGGCTGCCG 300
DB 428 GGAATTAACCAAGCCAGCGCTGCTGATGCGAGATGCTGCGCAAAAAGGCTGCCG 487
QY 301 CAGTCTGTCAGGTGCGGATTAATGCTGATGTCGACGAGCTTACCCGTAACAGCGCA 360
DB 488 CAGTCTGTCAGGTGCGGATTAATGCTGATGTCGACGAGCTTACCCGTAACAGCGCA 547
QY 361 GCTTCGCGCGCGCGCTGCGACCTGACCTGTCATTAACCTGACATACCAGGAGATACG 420
DB 548 GCTTCGCGCGCGCGCTGCGACCTGACCTGTCATTAACCTGACATACCAGGAGATACG 607
QY 421 TCCAGTCCCGATCCGTTATTTAATCCTTAATAAATGCGCTTGGCAATCGAATACCGG 480
DB 608 TCCAGTCCCGATCCGTTATTTAATCCTTAATAAATGCGCTTGGCAATCGAATACCGG 667
QY 481 AACGTACTGACGCGATCTCAGCAGGCGAGAGGCTCAATTGCTGACTTACCGGCAAT 540
DB 668 AACGTACTGACGCGATCTCAGCAGGCGAGAGGCTCAATTGCTGACTTACCGGCAAT 727
QY 541 CGGCAAAAGGGGTTTCGGAACTGGAACGGGTGCTTAATTTTCCGCAATCAAACTTGGC 600
DB 728 CGGCAAAAGGGGTTTCGGAACTGGAACGGGTGCTTAATTTTCCGCAATCAAACTTGGC 787
QY 601 CTTAAAGTGAAGAACAGAGCAAGAGCTTCAATTAACGAGGCAATTAACGAGACTC 660
DB 788 CTTAAAGTGAAGAACAGAGCAAGAGCTTCAATTAACGAGGCAATTAACGAGACTC 847
QY 661 AAGGTAGCGCGCAATGTCTCATTAACGGTGGTAACTGCAATCAATGCTGACG 720
DB 848 AAGGTAGCGCGCAATGTCTCATTAACGGTGGTAACTGCAATCAATGCTGACG 907
QY 721 GAGATATTTCTCTCCCAACAGCACAAGGAATGCGGAGCCGGGTGGGGAAGATCAC 780
DB 908 GAGATATTTCTCTCCCAACAGCACAAGGAATGCGGAGCCGGGTGGGGAAGATCAC 967
QY 781 GATTCAACACAGTGAACAACCTTGCTAAGTTGCAACGCGCAATTTTATTTGCTAACA 840
DB 968 GATTCAACACAGTGAACAACCTTGCTAAGTTGCTAAGTTGCTAAGTTGCTAACA 1027
QY 841 CGCAGCCGAGAGTTGCCCGCAGCGCGCAACCCGTTATTAATGATTTGATCAAGACG 900
DB 1028 CGCAGCCGAGAGTTGCCCGCAGCGCGCAACCCGTTATTAATGATTTGATCAAGACG 1087
QY 901 TTGAGCCGCCATCCACCGCAAAAAGGCGATGCTGATGACATTCACCACTTCAGTGTG 960
DB 1088 TTGAGCCGCCATCCACCGCAAAAAGGCGATGCTGATGACATTCACCACTTCAGTGTG 1147
QY 961 TTTATCGCCGACACGATATCTAATCTGCAAAATCTCGCGCGCGCATCTGAGCTCACTGG 1020
DB 1148 TTTATCGCCGACACGATATCTAATCTGCAAAATCTCGCGCGCGCATCTGAGCTCACTGG 1207
QY 1021 AGCTTCCCGGTACACCGGATTAACACGCCGCCGAGGTGTAATCTGCTTTTAAAGCTGG 1267

Db 1208 AGCTTCCCGGTGACCGGATTAACAAGCCGCAAGTGTGAATGTGTGTTTGAAGCTGG 1267
Qy 1081 CGTCGGCTAAGGATTAACAGCAATGATTCAGGTTTCCTGTGCTTCCAGACTTTACG 1140
Db 1268 CCGCGGCTAAGGATTAACAGCAATGATTCAGGTTTCCTGTGCTTCCAGACTTTACG 1327
Qy 1141 CAGATGCGGTAAACCGCGCTGTCAATTAATACGCGCGCGGAGAGGTGAACAGTAC 1200
Db 1328 CAGATGCGGTAAACCGCGCTGTCAATTAATACGCGCGCGGAGAGGTGAACAGTAC 1387
Qy 1201 CTGGCAGATGTAAGAGGAAATGCGCAGGCGATGTTCGTGACAGTTTACGCA 1260
Db 1388 CTGGCAGATGTAAGAGGAAATGCGCAGGCGATGTTCGTGACAGTTTACGCA 1447
Qy 1261 ATCGTGAATGAAGCAGCATACCGCGCTTGCACTTTGTA 1299
Db 1448 ATCGTGAATGAAGCAGCATACCGCGCTTGCACTTTGTA 1486
RESULT 9
US-10-601-319-6
; Sequence 6, Application US/10601319
; Publication No. US20040091968A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay M.
; APPLICANT: Kretz, Keith A.
; APPLICANT: Gray, Kevin A.
; APPLICANT: Barton, Nelson Robert
; APPLICANT: Garrett, James B.
; APPLICANT: O'Donoghue, Eileen
; APPLICANT: Mathur, Eric J.
; TITLE OF INVENTION: RECOMBINANT PHYTASES AND METHODS OF MAKING
; FILE REFERENCE: 09010-029011
; CURRENT APPLICATION NUMBER: US/10/601,319
; PRIOR FILING DATE: 2003-06-20
; PRIOR APPLICATION NUMBER: US 09/866,379
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: US 09/580,515
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 09/318,528
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: US 09/291,931
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: US 09/259,214
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: US 08/910,798
; PRIOR FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1901
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 403
; OTHER INFORMATION: n = A,T,C or G
US-10-601-319-6
Query Match 99.2%; Score 1288.4; DB 17; Length 1901;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1292; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 308 GTGGGTGCTCAACCAAGGCGACGCAATGATGCAAGATGTCAACCCAGACCATGCGCA 367
Qy 181 ACCGCGCGGTAAACCTGGGTGGCTGACACCGCGCGGTGGAGCTAATGCCATATCC 240
Db 368 ACCGCGCGGTAAACCTGGGTGGCTGACACCGCGCGGTGGAGCTAATGCCATATCC 427
Qy 241 GGAATTAACCAAGCGACGCGTGTGGTACGAGATTCGTGGCGGAAAGAGGAGTCCG 300
Db 428 GGAATTAACCAAGCGCGTGTGGTACGAGATTCGTGGCGGAAAGAGGAGTCCG 487
Qy 301 CAGTGTGTCAGATGCGCATTAATGCTGATGCAAGACGTAACCCGTAAACAGCGAA 360
Db 488 CAGTGTGTCAGATGCGCATTAATGCTGATGCAAGACGTAACCCGTAAACAGCGAA 547
Qy 361 GCGTTCGCGCGCGGCTGCGACCTGCACTGTGCAATACCTGATACCCAGAGCATAG 420
Db 548 GCGTTCGCGCGCGGCTGCGACCTGCACTGTGCAATACCTGATACCCAGAGCATAG 607
Qy 421 TCCAGTCCCGATCCGTTATTTAATCCTTAATAAAGGCGTTGCGCACTGATACGCG 480
Db 608 TCCAGTCCCGATCCGTTATTTAATCCTTAATAAAGGCGTTGCGCACTGATACGCG 667
Qy 481 AACGTACTGACGCGATCCTCAGACAGGAGAGGCTCAATGCTGACTTTACCGGCAAT 540
Db 668 AACGTACTGACGCGATCCTCAGACAGGAGAGGCTCAATGCTGACTTTACCGGCAAT 727
Qy 541 CCGCAAAAGCGGCTTTCGCAACTGGAAGCGGTCTTAATTTTCGCAATCAAACTGTGC 600
Db 728 CCGCAAAAGCGGCTTTCGCAACTGGAAGCGGTCTTAATTTTCGCAATCAAACTGTGC 787
Qy 601 CTTAAACGTGAAGAAAGAGACCAAGCTGTCAATTAACGAGCATTAACATCGAATC 660
Db 788 CTTAAACGTGAAGAAAGAGACCAAGCTGTCAATTAACGAGCATTAACATCGAATC 847
Qy 661 AAGGTAGCGCGCAATGCTCTTAATTAACCGGTGCGGTAAAGCTTCGATCAATGCTGACG 720
Db 848 AAGGTAGCGCGCAATGCTCTTAATTAACCGGTGCGGTAAAGCTTCGATCAATGCTGACG 907
Qy 721 GAGATATTTCTCTGCAACAGCAAGGAAATCCGAGCCGCGGTGGGAGAGATCAAC 780
Db 908 GAGATATTTCTCTGCAACAGCAAGGAAATCCGAGCCGCGGTGGGAGAGATCAAC 967
Qy 781 GATTCAACCAAGTGAACCTTGTGAATGTCATTAACGCGAATTTATTTGCTACAA 840
Db 968 GATTCAACCAAGTGAACCTTGTGAATGTCATTAACGCGAATTTATTTGCTACAA 1027
Qy 841 CGCAGCGCAGAGGTGCGCGCAGCGCCGCAACCCGTTAATTAATTAATCAAGACGCG 900
Db 1028 CGCAGCGCAGAGGTGCGCGCAGCGCCGCAACCCGTTAATTAATTAATCAAGACGCG 1087
Qy 901 TTGACGCCCATCAACCGCAAAAGCGGTATGTGTGACATTAACCACTTCACTGCTG 960
Db 1088 TTGACGCCCATCAACCGCAAAAGCGGTATGTGTGACATTAACCACTTCACTGCTG 1147
Qy 961 TTTATGCGCGGACACGATTAATCTGCAAAATCTGGCGCGGCACTGAGCTCAACTGG 1020
Db 1148 TTTATGCGCGGACACGATTAATCTGCAAAATCTGGCGCGGCACTGAGCTCAACTGG 1207
Qy 1021 AGCGTTCCCGGTGAGCGGATTAACGCGCGCAGGTGTGAATCTGCTGTTGAACGCTGG 1080
Db 1208 AGCGTTCCCGGTGAGCGGATTAACGCGCGCAGGTGTGAATCTGCTGTTGAACGCTGG 1267
Qy 1081 CGTCGGTAAAGCATTAACAGCAGTGTGATTCAGTGTTCCTTCAGACTTTACAG 1140
Db 1268 CGTCGGTAAAGCATTAACAGCAGTGTGATTCAGTGTTCCTTCAGACTTTACAG 1327
Qy 1141 CAGATGCGGTAAACCGCGCTGTCAATTAATACCGCGCGGAGAGGTGAACAGTAC 1200
Db 1328 CAGATGCGGTAAACCGCGCTGTCAATTAATACCGCGCGGAGAGGTGAACAGTAC 1387
Qy 1201 CTGGCAGATGTAAGAGGAAATGCGCAGGCGATGTTCGTGACAGTTTACGCA 1260

Db 1388 CTGGCAGGATGTGAGAGGAGGAATGCGAGGCGCATGTGTGTCGACAGTTTACGAA 1447
Qy 1261 ATCTGTAATGAAGCAGCATACCCCTTGCACTTTGTAA 1299
Db 1448 ATCTGTAATGAAGCAGCATACCCGCTGCACTTTGTAA 1486

RESULT 10
US-09-777-566A-1
; Sequence 1, Application US/09777566A
; Patent No. US20010055788A1
; GENERAL INFORMATION:
; APPLICANT: DIVERSA CORPORATION
; APPLICANT: SHORT, Jay
; APPLICANT: KRETZ, Keith
; TITLE OF INVENTION: RECOMBINANT BACTERIAL PHYTASES AND USES THEREOF
; FILE REFERENCE: DIVER1370-6
; CURRENT APPLICATION NUMBER: US/09/777,566A
; PRIOR FILING DATE: 2001-06-11
; PRIOR APPLICATION NUMBER: US 09/318,528
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: US 09/291,931
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: US 09/259,214
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: US 08/910,798
; PRIOR FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 1
; LENGTH: 1323
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1320)
; NAME/KEY: misc feature
; LOCATION: (1)..(1323)
; OTHER INFORMATION: n is any nucleotide
US-09-777-566A-1

Query Match 98.7%; Score 1282.2; DB 9; Length 1323;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 1287; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1 ATGAAAGCATTTAATCCCATTTTATCTCTTGATTCGTTAAACCCCGCAATCTGA 60
Db 1 ATGAAAGCATTTAATCCCATTTTATCTCTTGATTCGTTAAACCCCGCAATCTGA 60
Qy 61 TTGCGTCAGAGTGAAGCGGAGCTGGAAGCTGGAAGTGTGTGATGTTGTCAGTGTCAATGT 120
Db 61 TTGCGTCAGAGTGAAGCGGAGCTGGAAGCTGGAAGTGTGTGATGTTGTCAGTGTCAATGT 120
Qy 121 GTGCGTGTCTCAACCAAGGCGACGCAACTGATGACAGATGTCAACCCGACAGCGATGCCA 180
Db 121 GTGCGTGTCTCAACCAAGGCGACGCAACTGATGACAGATGTCAACCCGACAGCGATGCCA 180
Qy 181 ACCTGACCGGTAAACCTGGGTTGGCTGACACCGCGCGGTGTGAGCTTAATCGCTATCTC 240
Db 181 ACCTGACCGGTAAACCTGGGTTGGCTGACACCGCGCGGTGTGAGCTTAATCGCTATCTC 240
Qy 241 GGCATTTACCAAGCCAGCGCTCTGTGAGCCGAGATTGTGCGGAAAAAGGCGTGC 300
Db 241 GGCATTTACCAAGCCAGCGCTCTGTGAGCCGAGATTGTGCGGAAAAAGGCGTGC 300
Qy 301 CAGTCTGTGAGTCCGATTTATTTGCTGATGTGACGAGCGTACCCGTAACAGGCGAA 360
Db 301 CAGTCTGTGAGTCCGATTTATTTGCTGATGTGACGAGCGTACCCGTAACAGGCGAA 360
Qy 361 GCGTTGCGCGCGCGGCTGCGACCTGACTGTGCAATTAACCTGACATACCCGAGGAGATAG 420
Db 361 GCGTTGCGCGCGCGGCTGCGACCTGACTGTGCAATTAACCTGACATACCCGAGGAGATAG 420

Qy 421 TCCAGTCCCATCCGTTATTTAATCTCTAATAAACTGCGCTTTGCCAATGGAATACGCG 480
Db 421 TCCAGTCCCATCCGTTATTTAATCTCTAATAAACTGCGCTTTGCCAATGGAATACGCG 480
Qy 481 AACGTACTGACGGATCTCTAGCAGGGCAGAGAGGTCATTTGCTGACTTTACCGGGCAT 540
Db 481 AACGTACTGACGGATCTCTAGCAGGGCAGAGAGGTCATTTGCTGACTTTACCGGGCAT 540
Qy 541 CGGCAACCGGCTTTCCGGAATCTGAAACGGGTGCTTAATTTCCGCAATCAACTTGTGC 600
Db 541 CGGCAACCGGCTTTCCGGAATCTGAAACGGGTGCTTAATTTCCGCAATCAACTTGTGC 600
Qy 601 CTTAAACGTGAAACAGAGCAAAAGCTTTCAATTAACGAGGATTAACATCGAACTC 660
Db 601 CTTAAACGTGAAACAGAGCAAAAGCTTTCAATTAACGAGGATTAACATCGAACTC 660
Qy 661 AAGGTAGCGCGCAATGTCTCATTTAACCGGTGCGGTAGCTTCGATCAATGCTGACG 720
Db 661 AAGGTAGCGCGCAATGTCTCATTTAACCGGTGCGGTAGCTTCGATCAATGCTGACG 720
Qy 721 GAGATATTTCTCTGCAACAGCAAGGGAATGCGAGCGGGGTGGGAAAGATCAC 780
Db 721 GAGATATTTCTCTGCAACAGCAAGGGAATGCGAGCGGGGTGGGAAAGATCAC 780
Qy 781 GATTCAACACAGTGAACACCTTGCTAAGTTTGCAATAACGCGCAATTTATTTGCTACA 840
Db 781 GATTCAACACAGTGAACACCTTGCTAAGTTTGCAATAACGCGCAATTTATTTGCTACA 840
Qy 841 CGCAGCGCAGAGTTTCCCGCAGCGCGCGCACCCCGTTATTAATTTGATGAACAGCG 900
Db 841 CGCAGCGCAGAGTTTCCCGCAGCGCGCGCACCCCGTTATTAATTTGATGAACAGCG 900
Qy 901 TTGACGCCCATTCACCGCAAAAACGCGCTATGTGATGACATTAACCATTCAGTACTG 960
Db 901 TTGACGCCCATTCACCGCAAAAACGCGCTATGTGATGACATTAACCATTCAGTACTG 960
Qy 961 TTTATTCGCGGACAGATACTAATCTGCAAAATCTGCGCGCGCACTGAGCTCACTG 1020
Db 961 TTTATTCGCGGACAGATACTAATCTGCAAAATCTGCGCGCGCACTGAGCTCACTG 1020
Qy 1021 ACGCTTCCCGGTCAAGCGGATTAACGCGCGCAAGTGTGAACTGTGTTGAAAGCTGCG 1080
Db 1021 ACGCTTCCCGGTCAAGCGGATTAACGCGCGCAAGTGTGAACTGTGTTGAAAGCTGCG 1080
Qy 1081 CGTGGCTTAAGGATTAACAGCGAGTGAATTCAGTTTGCTGCTTCCAGCTTTACAG 1140
Db 1081 CGTGGCTTAAGGATTAACAGCGAGTGAATTCAGTTTGCTGCTTCCAGCTTTACAG 1140
Qy 1141 CAGATGCGGTAAACGCGCTGTCAATTAATACGCGCGCGGAGAGTGAACCTGACC 1200
Db 1141 CAGATGCGGTAAACGCGCTGTCAATTAATACGCGCGCGGAGAGTGAACCTGACC 1200
Qy 1201 CTGGCAGATGTGAAGAGCGAAATCGCAGGCGATGTGTTGTTGCGAGTTTACGAA 1260
Db 1201 CTGGCAGATGTGAAGAGCGAAATCGCAGGCGATGTGTTGTTGCGAGTTTACGAA 1260
Qy 1261 ATCTGTAATGAAGCAGCATACCCGCTTGCAGTTTG 1296
Db 1261 ATCTGTAATGAAGCAGCATACCCGCTTGCAGTTTG 1296

RESULT 11
US-09-866-379-1
; Sequence 1, Application US/09866379
; Patent No. US20020136754A1
; GENERAL INFORMATION:
; APPLICANT: DIVERSA CORPORATION
; APPLICANT: SHORT, Jay
; APPLICANT: KRETZ, Keith
; APPLICANT: GRAY, Kevin
; APPLICANT: BARTON, Nelson
; APPLICANT: GARRETT, James
; APPLICANT: O'DONOGHUE, Eileen

```

; TITLE OF INVENTION: RECOMBINANT BACTERIAL PHYTASES AND USES THEREOF
; FILE REFERENCE: DIVER1370-7
; CURRENT APPLICATION NUMBER: US/09/866,379
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: US 09/580,515
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 09/318,528
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: US 09/291,931
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: US 09/259,214
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: US 08/910,798
; PRIOR FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1323
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(1323)
; OTHER INFORMATION: n is any nucleotide
; NAME/KEY: CDS
; LOCATION: (1)..(1323)
; OTHER INFORMATION:
; US-09-866-379-1

```

```

Query Match      98.7%; Score 1282.2; DB 9; Length 1323;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 1287; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

```

```

QY 1 ATGAAGCCATCTTAATCCATTTTATCTCTCTGATTCGGTTAAACCCGCAATCTGCA 60
DB 1 ATGAAGCCATCTTAATCCATTTTATCTCTCTGATTCGGTTAAACCCGCAATCTGCA 60
QY 61 TTGCTCAGAGTAGAGCCGAGCTGAAGCTGAAAGTGTGTGATTTGTCATGCTCATGCT 120
DB 61 TTGCTCAGAGTAGAGCCGAGCTGAAGCTGAAAGTGTGTGATTTGTCATGCTCATGCT 120
QY 121 GTGCGTGCTTCAACCAAGGCCACGCAACTGATGAGATGTCAACCCGAGACGATGGCCA 180
DB 121 GTGCGTGCTTCAACCAAGGCCACGCAACTGATGAGATGTCAACCCGAGACGATGGCCA 180
QY 181 ACCTGCGCGGTAAACCTGGGTTGAGCTGACACCGCGCGGTGATGAGCTAATCGCTATCTC 240
DB 181 ACCTGCGCGGTAAACCTGGGTTGAGCTGACACCGCGCGGTGATGAGCTAATCGCTATCTC 240
QY 241 GGAATTATCCCAACGCCAGCGTGTGTGAGCCGAGCGGATTTGCTGGGAAAAAGGCTGCCG 300
DB 241 GGAATTATCCCAACGCCAGCGTGTGTGAGCCGAGCGGATTTGCTGGGAAAAAGGCTGCCG 300
QY 301 CAGTCTGATCAGATTCGGATTTATTTGCTGATGTGACAGAGGTACCCCTAAACAGGGGAA 360
DB 301 CAGTCTGATCAGATTCGGATTTATTTGCTGATGTGACAGAGGTACCCCTAAACAGGGGAA 360
QY 361 GCCTTGGCGCGCGGGCTGGACCTGACCTGTGCAATTAACGTTACCTCCAGGAGATACG 420
DB 361 GCCTTGGCGCGCGGGCTGGACCTGACCTGTGCAATTAACGTTACCTCCAGGAGATACG 420
QY 421 TCCAGTCCGATCCGTTATTTAATCCTCTAATAAACTGGCGTTTGGCAACTGATTAACGCG 480
DB 421 TCCAGTCCGATCCGTTATTTAATCCTCTAATAAACTGGCGTTTGGCAACTGATTAACGCG 480
QY 481 AAGGTGATCAGCGGATCTTCAAGAGGCGAGAGGTCATTTGCTGATTTTCCGGGCAAT 540
DB 481 AAGGTGATCAGCGGATCTTCAAGAGGCGAGAGGTCATTTGCTGATTTTCCGGGCAAT 540
QY 541 CGGCAAAAGGCGGTTTCCGGAATCGGAACGGGTGCTTAATTTTCCGCAATCAAACTTTGCG 600
DB 541 CGGCAAAAGGCGGTTTCCGGAATCGGAACGGGTGCTTAATTTTCCGCAATCAAACTTTGCG 600

```

```

QY 601 CTTAAACGTGAGAAACAGGACGAAGCTGTCTATTAAACGACGACATTAACATCGAATCTC 660
DB 601 CTTAAACGTGAGAAACAGGACGAAGCTGTCTATTAAACGACGACATTAACATCGAATCTC 660
QY 661 AAGGTGAGCGCCGACATGCTCTATTAAACCGGTGCGTAAAGCTTGGCATCAATGCTGACG 720
DB 661 AAGGTGAGCGCCGACATGCTCTATTAAACCGGTGCGTAAAGCTTGGCATCAATGCTGACG 720
QY 721 GAGATTTTCTCTCTGCAACACGACGAGAAAGCCGAGACCCGGGGGGGAAAGATCACG 780
DB 721 GAGATTTTCTCTCTGCAACACGACGAGAAAGCCGAGACCCGGGGGGGAAAGATCACG 780
QY 781 GATTACACACAGTGAACACCTTGTCTAAGTTTGCATTAACGCGCAATTTATTTGTACAA 840
DB 781 GATTACACACAGTGAACACCTTGTCTAAGTTTGCATTAACGCGCAATTTATTTGTACAA 840
QY 841 CGCAGCCGAGAGGTTGCCCCGACGCCGCCGCCGCAATTTAGATTGATCAAGACACG 900
DB 841 CGCAGCCGAGAGGTTGCCCCGACGCCGCCGCCGCAATTTAGATTGATCAAGACACG 900
QY 901 TTGACGCCCATTCACCGCAAAACAGGCGTATGTGATCAATTAACCACTTCAATCTG 960
DB 901 TTGACGCCCATTCACCGCAAAACAGGCGTATGTGATCAATTAACCACTTCAATCTG 960
QY 961 TTATGCGCGGACAGATCTAATCTGCAAAATCTGCGCGGCGCATGAGCTCAACTG 1020
DB 961 TTATGCGCGGACAGATCTAATCTGCAAAATCTGCGCGGCGCATGAGCTCAACTG 1020
QY 1021 ACGCTTCCCGGTACGCGGATTAACACGCGGCCGAGTGTGAACTGTGTTGAACCTGG 1080
DB 1021 ACGCTTCCCGGTACGCGGATTAACACGCGGCCGAGTGTGAACTGTGTTGAACCTGG 1080
QY 1081 CGTGCGCTAAGCGATTAACGCGAGTGTGATTCAGTTCGTTCTTCAGACTTTACAG 1140
DB 1081 CGTGCGCTAAGCGATTAACGCGAGTGTGATTCAGTTCGTTCTTCAGACTTTACAG 1140
QY 1141 CAGATCGGTATTAACGCGCGTGTCAATTAATACGCGCGCGAGAGGTGAACGTACG 1200
DB 1141 CAGATCGGTATTAACGCGCGTGTCAATTAATACGCGCGCGAGAGGTGAACGTACG 1200
QY 1201 CTGCGAGATGTGAAGAGCAAAATGCGCAGGCGATGTGTTGCTTGGCAGGTTTACGCA 1260
DB 1201 CTGCGAGATGTGAAGAGCAAAATGCGCAGGCGATGTGTTGCTTGGCAGGTTTACGCA 1260
QY 1261 ATCGTGAATGAACAGCATATCCGCGGTGCAGTTTG 1296
DB 1261 ATCGTGAATGAACAGCATATCCGCGGTGCAGTTTG 1296

```

```

RESULT 12
US-10-034-985-1
; Sequence 1, Application US/10034985
; Publication No. US20030049815A1
; GENERAL INFORMATION:
; APPLICANT: Kretz, Keith
; TITLE OF INVENTION: NOVEL PHYTASE
; FILE REFERENCE: 09010/029003
; CURRENT APPLICATION NUMBER: US/10/034,985
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: US/09/580,515
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: 09/291,931
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: 08/910,798
; PRIOR FILING DATE: 1997-08-13
; PRIOR APPLICATION NUMBER: 09/259,214
; PRIOR FILING DATE: 1999-03-01
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1323
; TYPE: DNA
; ORGANISM: Escherichia coli

```

```
FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1320)
; NAME/KEY: misc_feature
; LOCATION: (1)...(1323)
; OTHER INFORMATION: n = A,T,C or G
US-10-034-985-1

Query Match      98.7%; Score 1282.2; DB 14; Length 1323;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 1287; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 ATGAAGCATCTTAATCCATTTTATCTCTTGATTCGCTTAACCCGCAATCTGCA 60
DB 1 ATGAAGCATCTTAATCCATTTTATCTCTTGATTCGCTTAACCCGCAATCTGCA 60
QY 61 TTGCTCAGAGTGAGCCGAGCTGGAAGTGAGTGATTTGTCAGTGTGATGAT 120
DB 61 TTGCTCAGAGTGAGCCGAGCTGGAAGTGAGTGATTTGTCAGTGTGATGAT 120
QY 121 GTGCTGCTCCAAACCAAGCCGACGCACTGATGCAAGATGTACCCCGACGATGGCCA 180
DB 121 GTGCTGCTCCAAACCAAGCCGACGCACTGATGCAAGATGTACCCCGACGATGGCCA 180
QY 121 GTGCTGCTCCAAACCAAGCCGACGCACTGATGCAAGATGTACCCCGACGATGGCCA 180
DB 121 GTGCTGCTCCAAACCAAGCCGACGCACTGATGCAAGATGTACCCCGACGATGGCCA 180
QY 181 ACCCTGCGGTAAACTGGGTTGGCTGACACCGCGGTGTGAGCTAATCGCTATCTC 240
DB 181 ACCCTGCGGTAAACTGGGTTGGCTGACACCGCGGTGTGAGCTAATCGCTATCTC 240
QY 241 GGACATTTACCAACGCCAGCGTCTGTAGCCGACGAAATTGCTGGCGAAAAAGGCTGCCG 300
DB 241 GGACATTTACCAACGCCAGCGTCTGTAGCCGACGAAATTGCTGGCGAAAAAGGCTGCCG 300
QY 301 CAGTCTGTGACAGTCCGCAATTATTTGCTGATGTCAGACGCTTAACCCGTAACAGGCGAA 360
DB 301 CAGTCTGTGACAGTCCGCAATTATTTGCTGATGTCAGACGCTTAACCCGTAACAGGCGAA 360
QY 361 GCCTTGCGCGCGGGGTGACCTGCTGCAATACCGTCAATCCGACGACGATAG 420
DB 361 GCCTTGCGCGCGGGGTGACCTGCTGCAATACCGTCAATCCGACGACGATAG 420
QY 421 TCCAGTCCCGATCCGTTATTTAATCTCTTAAAACTGGGCTTGGCAATGATTAACGCG 480
DB 421 TCCAGTCCCGATCCGTTATTTAATCTCTTAAAACTGGGCTTGGCAATGATTAACGCG 480
QY 481 AACGTACTGACGCGCATCTCTCAGACGAGGAGGCTCAATTCCTTAACCCGCGCAT 540
DB 481 AACGTACTGACGCGCATCTCTCAGACGAGGAGGCTCAATTCCTTAACCCGCGCAT 540
QY 541 CGGCAAAACGCGCTTTGCGCAACTGAAACGAGGTCTTAATTTTCCGCAATCAAACTTGTC 600
DB 541 CGGCAAAACGCGCTTTGCGCAACTGAAACGAGGTCTTAATTTTCCGCAATCAAACTTGTC 600
QY 601 CTTAAACGTGAGAAACAGACGAAAGCTGTTCAATTAACGAGCATTTACCATCGAACTC 660
DB 601 CTTAAACGTGAGAAACAGACGAAAGCTGTTCAATTAACGAGCATTTACCATCGAACTC 660
QY 661 AAGGTAGGCGCCGACATGCTCTTAACCGGTGCGGTAACTCCGCAATCAATCGTAGG 720
DB 661 AAGGTAGGCGCCGACATGCTCTTAACCGGTGCGGTAACTCCGCAATCAATCGTAGG 720
QY 721 GAGATATTTCTCTGCAACAGACAGGAAATCCGAGCGGGGTGGGGAAGATCACC 780
DB 721 GAGATATTTCTCTGCAACAGACAGGAAATCCGAGCGGGGTGGGGAAGATCACC 780
QY 781 GATTCAACACAGTGAACAACCTTGCTAAAGTTTGCAATACCGCAATTTTATTTGCTACAA 840
DB 781 GATTCAACACAGTGAACAACCTTGCTAAAGTTTGCAATACCGCAATTTTATTTGCTACAA 840
QY 841 CGACGCGCAGAGGTGGCGCGACCGCGGCAACCCCGTATTAATTTGATGACAGACGG 900
DB 841 CGACGCGCAGAGGTGGCGCGACCGCGGCAACCCCGTATTAATTTGATGACAGACGG 900
QY 901 TTGACGCGCATCCACGCAAAAACAGGCGTATGTGTGACATTACCACTTCAGTGTG 960
```

```
DB 901 TTGACGCGCATCCACGCAAAAACAGGCGTATGTGTGACATTACCACTTCAGTGTG 960
QY 961 TTATATCCCGGACACGATACTATCTGCAAAATCTCGCGCGCGGACATGAGCTCACTCG 1020
DB 961 TTATATCCCGGACACGATACTATCTGCAAAATCTCGCGCGCGGACATGAGCTCACTCG 1020
QY 1021 ACGCTTCCCGGTGACCGCGATTAACACGCGCGCAGAGGTGTAATGAGTTTGAACGCTGG 1080
DB 1021 ACGCTTCCCGGTGACCGCGATTAACACGCGCGCAGAGGTGTAATGAGTTTGAACGCTGG 1080
QY 1081 CGTGGCTTAACCGATTAACGCGCAAGTGAATTCAGTTTCTGCTGTCTTCCAGACTTTACAG 1140
DB 1081 CGTGGCTTAACCGATTAACGCGCAAGTGAATTCAGTTTCTGCTGTCTTCCAGACTTTACAG 1140
QY 1141 CAGATGCTGATTAACCGCGCTGCTGCTTAATACCGCGCGGAGAGGTGAACCTGACC 1200
DB 1141 CAGATGCTGATTAACCGCGCTGCTGCTTAATACCGCGCGGAGAGGTGAACCTGACC 1200
QY 1201 CTGGCAGAGTGTGAAGACGAAATGCGCAGGSCATGTTGCTTGGCAGTTTTACGCA 1260
DB 1201 CTGGCAGAGTGTGAAGACGAAATGCGCAGGSCATGTTGCTTGGCAGTTTTACGCA 1260
QY 1261 ATCGTAATGAAGACGCAATACCGCTTGCAATTG 1296
DB 1261 ATCGTAATGAAGACGCAATACCGCTTGCAATTG 1296
```

RESULT 13

```
US-10-430-356-1
; Sequence 1, Application US/10430356
; Publication No. US20030232041A1
; GENERAL INFORMATION:
; APPLICANT: Kretz, Keith
; TITLE OF INVENTION: NOVEL PHYTASE
; FILE REFERENCE: 09010/029003
; CURRENT APPLICATION NUMBER: US/10/430,356
; CURRENT FILING DATE: 2003-05-05
; PRIOR APPLICATION NUMBER: US/09/580,515
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: 09/318,528
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: 09/291,931
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: 08/910,798
; PRIOR FILING DATE: 1997-08-13
; PRIOR APPLICATION NUMBER: 09/259,214
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1323
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1320)
; NAME/KEY: misc_feature
; LOCATION: (1)...(1323)
; OTHER INFORMATION: n = A,T,C or G
US-10-430-356-1
```

```
Query Match      98.7%; Score 1282.2; DB 17; Length 1323;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 1287; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 ATGAAGCATCTTAATCCATTTTATCTCTTGATTCGCTTAACCCGCAATCTGCA 60
DB 1 ATGAAGCATCTTAATCCATTTTATCTCTTGATTCGCTTAACCCGCAATCTGCA 60
QY 61 TTGCTCAGAGTGAGCCGAGCTGGAAGTGAGTGATTTGTCAGTGTGATGAT 120
```


Db 61 TTGCTCAGAGTGAAGCCGAGCTGAAGCTGAAAGTGTGTGATTTGTCACTGTCATGCT 120
Qy 121 GTGCTGTCTCAACCAAGGCCAAGCCAGCACTGATGAGAGATGTGACCCCGACAGCATGCGCA 180
Db 121 GTGGGTGTCTCAACCAAGGCCAAGCCAGCACTGATGAGAGATGTGACCCCGACAGCATGCGCA 180
Qy 181 ACCTGGCCGGTAAACTGGGTGGTGTGACACCGCGGGTGTGTGAGTAAATCCCTATCTC 240
Db 181 ACCTGGCCGGTAAACTGGGTGGTGTGACACCGCGGGTGTGTGAGTAAATCCCTATCTC 240
Qy 241 GGACATTAACCAAGCCAGCGCTGTGTGAGCGCAAGATTCGTGCGCAAAAAAGGGCTGCGCG 300
Db 241 GGACATTAACCAAGCCAGCGCTGTGTGAGCGCAAGATTCGTGCGCAAAAAAGGGCTGCGCG 300
Qy 301 CAGTCTGTGTCAGGTGCGCATTAATGCTGATGTCGAGCGATACCCGTAAACAGCGCA 360
Db 301 CAGTCTGTGTCAGGTGCGCATTAATGCTGATGTCGAGCGATACCCGTAAACAGCGCA 360
Qy 361 GCGTTCGCGCGCGGGCTGGCACTGACTGTGCAATACCGTACATCCAGGCGATACG 420
Db 361 GCGTTCGCGCGCGGGCTGGCACTGACTGTGCAATACCGTACATCCAGGCGATACG 420
Qy 421 TCCAGTCCGATCCGTTATTTAATCTCTAATAAACTGGGTTGCCAATGATTAACGCG 480
Db 421 TCCAGTCCGATCCGTTATTTAATCTCTAATAAACTGGGTTGCCAATGATTAACGCG 480
Qy 481 AACGTACTGACGCGATCTCTACAGAGGAGAGAGGATTCATTTGCTGACTTACCGGGCAT 540
Db 481 AACGTACTGACGCGATCTCTACAGAGGAGAGAGGATTCATTTGCTGACTTACCGGGCAT 540
Qy 541 CCGCAAAACGCGCTTTGCGCAACTGGAACCGGCTCTTAATTTTCCGAACTCAAACTTGTC 600
Db 541 CCGCAAAACGCGCTTTGCGCAACTGGAACCGGCTCTTAATTTTCCGAACTCAAACTTGTC 600
Qy 601 CTTAAACGTGAAGAAAGAGAGCGAAGCTGTTCAATTAACGAGGATTAACCATCGAATC 660
Db 601 CTTAAACGTGAAGAAAGAGAGCGAAGCTGTTCAATTAACGAGGATTAACCATCGAATC 660
Qy 661 AAGGTAGGCGCCGACATGCTCTATTAAACGGTGGTGAAGCCTGCGATCAATGCTGACG 720
Db 661 AAGGTAGGCGCCGACATGCTCTATTAAACGGTGGTGAAGCCTGCGATCAATGCTGACG 720
Qy 721 GAGATATTTCTCTCTGCAACAGCAACAGGAATCCCGAGCGGGGTGGGGAAGATCAC 780
Db 721 GAGATATTTCTCTCTGCAACAGCAACAGGAATCCCGAGCGGGGTGGGGAAGATCAC 780
Qy 781 GATTTCACACAGTGAACACCTTGTTGATTAAGCGGCAATTTTATTTGCTCA 840
Db 781 GATTTCACACAGTGAACACCTTGTTGATTAAGCGGCAATTTTATTTGCTCA 840
Qy 841 CGCAGCCAGAGGTTGGCCGACCGCGCCACCCCGTTATAGATTGATCAAGACAGCG 900
Db 841 CGCAGCCAGAGGTTGGCCGACCGCGCCACCCCGTTATAGATTGATCAAGACAGCG 900
Qy 901 TTGACGCCCATTCACCGCAAAAACAGGCGTATGTGTGACATTAACCATTTGACTG 960
Db 901 TTGACGCCCATTCACCGCAAAAACAGGCGTATGTGTGACATTAACCATTTGACTG 960
Qy 961 TTTATGCGCGGAGACGATTAATCTGCGCAATCTCGCGCGGCGCATGAGCTCACTGG 1020
Db 961 TTTATGCGCGGAGACGATTAATCTGCGCAATCTCGCGCGGCGCATGAGCTCACTGG 1020
Qy 1021 ACCTTCGCGGATCGCGGATTAACCGCGCGAGAGTGTGAATCTGTTTGAACGCTGG 1080
Db 1021 ACCTTCGCGGATCGCGGATTAACCGCGCGAGAGTGTGAATCTGTTTGAACGCTGG 1080
Qy 1081 CGTGGCTTAAGCATTAACAGCAGTGAATTCAGGTTTCTGCTGCTTCAGACTTTTACAG 1140
Db 1081 CGTGGCTTAAGCATTAACAGCAGTGAATTCAGGTTTCTGCTGCTTCAGACTTTTACAG 1140
Qy 1141 CAGATGCGTAAATAAGCGCGCTGTCAATTAATGCGCGCGCGAGAGGATGAATCTGACC 1200
Db 1141 CAGATGCGTAAATAAGCGCGCTGTCAATTAATGCGCGCGCGAGAGGATGAATCTGACC 1200

Qy 1201 CTTGCGAGATGTGAAGAGCGAAATGCGAGGCGATGTGTTGGTGGAGTTTACGCA 1260
Db 1201 CTTGCGAGATGTGAAGAGCGAAATGCGAGGCGATGTGTTGGTGGAGTTTACGCA 1260
Qy 1261 ATCGTGAATGAAGCAGCATACCGCTTGCACTTTG 1296
Db 1261 ATCGTGAATGAAGCAGCATACCGCGTGCAGTTTG 1296

RESULT 14
US-10-601-319-1
Sequence 1, Application US/10601319
Publication No. US20040091968A1
GENERAL INFORMATION:
APPLICANT: Short, Jay M.
APPLICANT: Kretz, Keith A.
APPLICANT: Gray, Kevin A.
APPLICANT: Barton, Nelson Robert
APPLICANT: Garrett, James B.
APPLICANT: O'Donoghue, Eileen
APPLICANT: Mathur, Eric J.
TITLE OF INVENTION: RECOMBINANT PHYLASES AND METHODS OF MAKING
FILE REFERENCE: 09010-029011
CURRENT APPLICATION NUMBER: US/10/601,319
PRIORITY FILING DATE: 2003-06-20
PRIORITY FILING DATE: 2001-05-24
PRIORITY FILING DATE: 2001-05-24
PRIORITY FILING DATE: 2000-05-25
PRIORITY FILING DATE: 1999-05-25
PRIORITY FILING DATE: 1999-04-13
PRIORITY FILING DATE: 1999-04-13
PRIORITY FILING DATE: 1999-03-01
PRIORITY FILING DATE: 1997-08-13
NUMBER OF SEQ ID NOS: 10
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 1323
TYPE: DNA
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(1320)
FEATURE:
NAME/KEY: misc_feature
LOCATION: 216
OTHER INFORMATION: n = A,T,C or G
US-10-601-319-1

Query Match 98.7%; Score 1282.2; DB 17; Length 1323;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 1287; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1 ATGAAGCCATCTTAATCCCATTTTATCTCTGATTCGTTAACCCGCAATCTGCA 60
Db 1 ATGAAGCCATCTTAATCCCATTTTATCTCTGATTCGTTAACCCGCAATCTGCA 60
Qy 61 TTGCTCAGAGTGAAGCCGAGCTGAAGCTGGAAAGTGTGTGATTTGTCATGCTGATGT 120
Db 61 TTGCTCAGAGTGAAGCCGAGCTGAAGCTGGAAAGTGTGTGATTTGTCATGCTGATGT 120
Qy 121 GTGGGTGTCTCAACCAAGGCCAAGCCAGCACTGATGAGAGATGTGACCCCGACAGCATGCGCA 180
Db 121 GTGGGTGTCTCAACCAAGGCCAAGCCAGCACTGATGAGAGATGTGACCCCGACAGCATGCGCA 180
Qy 181 ACCTGGCCGGTAAACTGGGTGGTGTGACACCGCGGGTGTGTGAGTAAATCCCTATCTC 240
Db 181 ACCTGGCCGGTAAACTGGGTGGTGTGACACCGCGGGTGTGTGAGTAAATCCCTATCTC 240

```

QY 241 GGAATTAACCAACGCGAGCTGTGGTAGCCGAGATTGCTGGCGAAGGAGGCTGCCG 300
DB 241 GGAATTAACCAACGCGAGCTGTGGTAGCCGAGATTGCTGGCGAAGGAGGCTGCCG 300
QY 301 CAGTCTGTGAGGTGCGGATTAATGCTGTGACGACGCTGATCCCGTAAACAGCGCA 360
DB 301 CAGTCTGTGAGGTGCGGATTAATGCTGTGACGACGCTGATCCCGTAAACAGCGCA 360
QY 361 GCTTGGCGCGCGGCTGGACCTGATCTGTGCAATTAACCGTACATACCGGACATACG 420
DB 361 GCTTGGCGCGCGGCTGGACCTGATCTGTGCAATTAACCGTACATACCGGACATACG 420
QY 421 TCCAGTCCCGATCCGTTATTTATCTCTAAACCTGGCGTTGGCACTGGATTAACGG 480
DB 421 TCCAGTCCCGATCCGTTATTTATCTCTAAACCTGGCGTTGGCACTGGATTAACGG 480
QY 481 AACGTGACTGACGCGATCTCTAGACGAGGAGGATCAATTGCTGACTTTACCGGGCAT 540
DB 481 AACGTGACTGACGCGATCTCTAGACGAGGAGGATCAATTGCTGACTTTACCGGGCAT 540
QY 541 CGGCAAAACGCGCTTTGCGCACTGGAACGCGGCTTAATTTTCCGCAATCAACTTGGC 600
DB 541 CGGCAAAACGCGCTTTGCGCACTGGAACGCGGCTTAATTTTCCGCAATCAACTTGGC 600
QY 601 CTTAAACGTGAAGAACAGGACGAAAGCTGTTCATTAAACGAGGCAATTAACATCGAACTC 660
DB 601 CTTAAACGTGAAGAACAGGACGAAAGCTGTTCATTAAACGAGGCAATTAACATCGAACTC 660
QY 661 AAGTGAGCGCGCAATGTCTCATTAACCGGTGCGGTAAGCTTCGATCAATGCTGACG 720
DB 661 AAGTGAGCGCGCAATGTCTCATTAACCGGTGCGGTAAGCTTCGATCAATGCTGACG 720
QY 721 GAGATTTTCTCTGGAACAGGAGGAAATCCGAGGCGGAGGAGGAGGAGATCACG 780
DB 721 GAGATTTTCTCTGGAACAGGAGGAAATCCGAGGCGGAGGAGGAGGAGATCACG 780
QY 781 GATTCAACCAAGTGAACACCTTGTCAATTTGCAATACGCGCAATTTATTTGCTACA 840
DB 781 GATTCAACCAAGTGAACACCTTGTCAATTTGCAATACGCGCAATTTATTTGCTACA 840
QY 841 CGCAGCGCAGAGGTTGCGCGACGCGGCAACCCCGTTATTAAGTTGATCAAGACGCG 900
DB 841 CGCAGCGCAGAGGTTGCGCGACGCGGCAACCCCGTTATTAAGTTGATCAAGACGCG 900
QY 901 TTGACCGCCCATTCACCGCAAAACAGGCGTATGTGACATTAACCTTCAGTACG 960
DB 901 TTGACCGCCCATTCACCGCAAAACAGGCGTATGTGACATTAACCTTCAGTACG 960
QY 961 TTTATCGCGGACACGATTAATCTGCAAAATCTGCGGCGGCACTGAGCTCACTG 1020
DB 961 TTTATCGCGGACACGATTAATCTGCAAAATCTGCGGCGGCACTGAGCTCACTG 1020
QY 1021 AGCGTTCCCGGTAGCGCGAATACGCGCGCGAGTGTGAATGTTGAAACCTG 1080
DB 1021 AGCGTTCCCGGTAGCGCGAATACGCGCGCGAGTGTGAATGTTGAAACCTG 1080
QY 1081 CGTGGCTAAGCGATTAACGCGCGGATTAAGTTGCTGCTTCAGACTTTACG 1140
DB 1081 CGTGGCTAAGCGATTAACGCGCGGATTAAGTTGCTGCTTCAGACTTTACG 1140
QY 1141 CAGATCGTGAATTAACGCGCTGTCAATTAATACGCGCGCGAGAGGTGAATGAC 1200
DB 1141 CAGATCGTGAATTAACGCGCTGTCAATTAATACGCGCGCGAGAGGTGAATGAC 1200
QY 1201 CTGGCAGAGTGAAGAGGAAATGCGCAGGCGATGTTGTTGCGAGTTTACGCA 1260
DB 1201 CTGGCAGAGTGAAGAGGAAATGCGCAGGCGATGTTGTTGCGAGTTTACGCA 1260
QY 1261 ATGCTGAATGAACGCGATTAACCGGCTTGAGTTG 1296
DB 1261 ATGCTGAATGAACGCGATTAACCGGCTTGAGTTG 1296

```

```

RESULT 15
US-10-156-660-1
; Sequence 1, Application US/10156660
; Publication No. US20030103958A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay M.
; APPLICANT: Kretz, Keith
; APPLICANT: Gray, Kevin A.
; APPLICANT: Barton, Nelson R.
; APPLICANT: Garrett, James B.
; APPLICANT: O'Donoghue, Bileen
; APPLICANT: Machur, Eric J.
; TITLE OF INVENTION: PHYLASES, NUCLEIC ACIDS ENCODING THEM
; FILE OF INVENTION: AND METHODS FOR MAKING AND USING THEM
; FILE REFERENCE: 09010-023007
; CURRENT APPLICATION NUMBER: US/10/156,660
; PRIOR FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: US 09/866,379
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: US 09/580,515
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 09/318,528
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: US 09/291,931
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: US 09/259,214
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: US 08/910,798
; PRIOR FILING DATE: 1997-08-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1308
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: modified phylase enzyme
; NAME/KEY: CDS
; LOCATION: (1)...(1308)
; US-10-156-660-1

Query Match          96.9%; Score 1259.2; DB 15; Length 1308;
Best Local Similarity 98.2%; Pred. No. 0;
Matches 1273; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 1 ATGAAAGCATTTTATCCATTTTATCTCTGATTCGTTAACCCTGCAATCTGCA 60
DB 1 ATGAAAGCATTTTATCCATTTTATCTCTGATTCGTTAACCCTGCAATCTGCA 60
QY 61 TTGCGTCAGAGTGAACCGGAGCTGAAGCTGAAGAGTGTGATTTGATGATGTCATGGT 120
DB 61 TTGCGTCAGAGTGAACCGGAGCTGAAGCTGAAGAGTGTGATTTGATGATGTCATGGT 120
QY 121 GTGCGTCTCCACCAAGGCGACGCACTGATGACGATGTCACCCGACGAGTGGCA 180
DB 121 GTGCGTCTCCACCAAGGCGACGCACTGATGACGATGTCACCCGACGAGTGGCA 180
QY 121 GTGCGTCTCCACCAAGGCGACGCACTGATGACGATGTCACCCGACGAGTGGCA 180
DB 121 GTGCGTCTCCACCAAGGCGACGCACTGATGACGATGTCACCCGACGAGTGGCA 180
QY 181 ACTGCGCGGTAAACTGGGTGGCTGACACCGCGGAGTGTGAGTGAATGCTATCTC 240
DB 181 ACTGCGCGGTAAACTGGGTGGCTGACACCGCGGAGTGTGAGTGAATGCTATCTC 240
QY 241 GGAATTAACCAACGCGAGCTGTGGTAGCCGAGATTGCTGGCGAAGGAGGCTGCCG 300
DB 241 GGAATTAACCAACGCGAGCTGTGGTAGCCGAGATTGCTGGCGAAGGAGGCTGCCG 300
QY 301 CAGTCTGTGAGGTGCGGATTAATGCTGTGACGACGCTGATCCCGTAAACAGCGCA 360
DB 301 CAGTCTGTGAGGTGCGGATTAATGCTGTGACGACGCTGATCCCGTAAACAGCGCA 360
QY 361 GCTTGGCGCGCGGCTGGACCTGATCTGTGCAATTAACCGTACATACCGGACATACG 420
DB 361 GCTTGGCGCGCGGCTGGACCTGATCTGTGCAATTAACCGTACATACCGGACATACG 420

```

Db 361 GCCTTGCCGCCGGGCTGGCACCTGACTGCAATTAACCGTACATTAACCCAGGAGATACG 420
QY 421 TCCAGTCCCGATCCGTTATTTAACTCTTAATAAACTGGGTTTGCCACTGATTAACGCG 480
Db 421 TCCAGTCCCGATCCGTTATTTAACTCTTAATAAACTGGGTTTGCCACTGATTAACGCG 480
QY 481 AACGTGACTGACGCGATCTCTCAGCAGGGGAGAGGGTCAATTGCTGACTTTTACCGGGCAT 540
Db 481 AACGTGACTGACGCGATCTCTCAGAGGGGAGAGGGTCAATTGCTGACTTTTACCGGGCAT 540
QY 541 CCGCAAAACGGCGTTGCGCACTGGAACGGGTGCTTAATTTTCGCAATCAAACTTGTC 600
Db 541 TATCAAAACGGCGTTGCGCACTGGAACGGGTGCTTAATTTTCGCAATCAAACTTGTC 600
QY 601 CTTAAACGTGAGAAACAGGACGAAAGCTGTTCATTAACGAGGCAATTACCATCGAACTC 660
Db 601 CTTAAACGTGAGAAACAGGACGAAAGCTGTTCATTAACGAGGCAATTACCATCGAACTC 660
QY 661 AAGGTAGCGCCGCAATGTCTTATTAACGGGTGCGGTAAAGCTTCGCAATCAATGCTGACG 720
Db 661 AAGGTAGCGCCGCAATGTCTTATTAACGGGTGCGGTAAAGCTTCGCAATCAATGCTGACG 720
QY 721 GAGATATTTCTCTGCAACAGCAAGGAAATGCCGAGCCGGGATGGGGAAGATCAC 780
Db 721 GAGATATTTCTCTGCAACAGCAAGGAAATGCCGAGCCGGGATGGGGAAGATCAC 780
QY 781 GATTACACCAAGTGAACACTTGTCTAAATTTGCATTAACGCGCAATTTATTTTGTACAA 840
Db 781 GATTACACCAAGTGAACACTTGTCTAAATTTGCATTAACGCGCAATTTATTTTGTACAA 840
QY 841 CGCAGCCGAGAGGTGGCCGCGAGCCGCGCAACCCGTTATTGAATTTGATCAAGACAGCG 900
Db 841 CGCAGCCGAGAGGTGGCCGCGAGCCGCGCAACCCGTTATTGAATTTGATCAAGACAGCG 900
QY 901 TTTACGCCCCATCAACGCAAAAACAGGCGTATGTGTGAATTAACCACTTCAGTCTG 960
Db 901 TTTACGCCCCATCAACGCAAAAACAGGCGTATGTGTGAATTAACCACTTCAGTCTG 960
QY 961 TTTATCGCCGAGACGATTAATCTGCAAAATCTCGGCGGCGCATGAGCTCACTG 1020
Db 961 TTTATCGCCGAGACGATTAATCTGCAAAATCTCGGCGGCGCATGAGCTCACTG 1020
QY 1021 ACGCTTCCCGGTACGCGGATTAACGCGCGCAGGTGTGAATCTGTGTTGAACGCTGG 1080
Db 1021 ACGCTTCCCGGTACGCGGATTAACGCGCGCAGGTGTGAATCTGTGTTGAACGCTGG 1080
QY 1081 CGTCCGCTAAGCATTAACGCCAGTGAATCAGGTTGCTGCTCAAGCTTTACAG 1140
Db 1081 CGTCCGCTAAGCATTAACGCCAGTGAATCAGGTTGCTGCTCAAGCTTTACAG 1140
QY 1141 CAGATGCGTATTAACGCCGCTGTCAATTAATACGCCGCCGAGAGGTGAACGTGACC 1200
Db 1141 CAGATGCGTATTAACGCCGCTGTCAATTAATACGCCGCCGAGAGGTGAACGTGACC 1200
QY 1201 CTGGCAGATGTGAAGAGCAAAATGCGAGGGCATGTGTTGCGAGAGTTTACGCA 1260
Db 1201 CTGGCAGATGTGAAGAGCAAAATGCGAGGGCATGTGTTGCGAGAGTTTACGCA 1260
QY 1261 ATGCTGAATGAAGACGATTAACCGCTTGAGGTTG 1296
Db 1261 ATGCTGAATGAAGACGATTAACCGCGTGAGTTG 1296

Search completed: February 10, 2005, 18:17:13
Job time : 956.54 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 22:02:33 ; Search time 5631.01 Seconds
(without alignments)
8780.927 Million cell updates/sec

Title: US-09-926-375b-7_COPY_12653_13951

Perfect score: 1299

Sequence: 1 atgaagacatcttaacc.....taccgcgtgcagttgttaa 1299

Scoring table: IDENTITY NUC

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST.*
1: gb_est1.*
2: gb_est2.*
3: gb_hic.*
4: gb_est3.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_gest1.*
9: gb_gest2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	614.4	47.3	616	BU074127	BU074127 BU074127
2	580.8	44.7	853	CL662734	CL662734 PRI0142b
3	536.2	41.3	746	CA093060	SCCCCL200
4	529	40.7	529	BU171370	SJABUC02
5	488.8	37.6	595	CF325092	CF325092 JMT1--05-
6	483.8	37.2	895	CN762997	ID0AA5DC
7	445.8	34.3	868	CN754382	ID0AA12D
8	441.8	34.0	864	CN757565	ID0AA1DF
9	386.2	29.7	393	BE520240	M1B12STM
10	383	29.5	383	AM036132	AM036132 EST274510
11	383	29.5	383	AM036134	AM036134 EST274510
12	353.4	27.2	531	BU171376	SJABUC01
13	304.6	23.4	354	AA545747	HBMSF1B4-
14	294.8	22.7	706	CN759004	ID0AA24B
15	256	19.7	707	BU619443	BU619443
16	216.8	16.7	320	BE436403	BE436403 EST407481
17	208	16.0	208	BM409016	EST583343
18	208	16.0	208	BM412806	EST587133
19	198.8	15.3	546	BU040796	BU040796
20	169.4	13.0	179	BE461872	EST413291
21	162.4	12.5	657	BG457132	NF100C05P
22	160.8	12.4	508	BU030203	BU030203
23	141.6	10.9	254	BE520241	M1B12XTM
24	103.4	8.0	769	CO386885	CO386885 AGENCOURT

C 25	54.6	4.2	821	9	CL665831	CL665831 PRI014d_E
C 26	54.6	4.2	842	7	CL665745	CL665745 PRI014c_E
C 27	43.8	3.4	392	7	H60613	H60613 YR4d12.B1
C 28	42.2	3.2	733	4	BI152280	BI152280 602917706
C 29	40.8	3.1	346	5	BY053637	BY053637 BY053637
C 30	40.8	3.1	350	5	BY041909	BY041909 BY041909
C 31	40.8	3.1	365	5	BY056036	BY056036 BY056036
C 32	40.8	3.1	389	6	CA566872	CA566872 K0406H12-
C 33	40.8	3.1	431	1	AI553154	AI553154 v107b12.Y
C 34	40.8	3.1	539	5	BX517075	BX517075 BX517075
C 35	40.8	3.1	602	2	BB610221	BB610221 BB610221
C 36	40.8	3.1	607	1	AA521521	AA521521 v107b12.T
C 37	40.8	3.1	632	6	BY738708	BY738708 BY738708
C 38	40.8	3.1	646	6	BY737657	BY737657 BY737657
C 39	40.8	3.1	2931	2	AK075739	AK075739 Mus muscu
C 40	39.2	3.0	379	2	BE289814	BE289814 601089743
C 41	39.2	3.0	411	6	CA880619	CA880619 K0984D03-
C 42	39.2	3.0	424	5	BY261762	BY261762 BY261762
C 43	39.2	3.0	434	5	BY238933	BY238933 BY238933
C 44	39.2	3.0	436	6	CA895377	CA895377 B0191P11-
C 45	39.2	3.0	485	6	CA888817	CA888817 B0148G10-

ALIGNMENTS

RESULT 1
BU074127/c 616 bp mRNA linear EST 29-SEP-2003
BU074127 NIBB Mochii normalized Xenopus tailbud library Xenopus
LOCUS
DEFINITION
laevis cDNA clone XLO90j06 5', mRNA sequence.
ACCESSION
BU074127 GI:17504316
VERSION
BU074127.1
KEYWORDS
EST.
SOURCE
Xenopus laevis (African clawed frog)
ORGANISM
Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
Xenopodinae; Xenopus; Xenopus.
REFERENCE
1 (bases 1 to 616)
Kikayama,A., Terasaka,C., Mochii,M., Ueno,N., Shn-i,T. and
Kohara,Y.
Expressed genes in X. laevis embryo
TITLE
Unpublished (2001)
JOURNAL
Contact: Tadao Shin-i
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
COMMENT
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp
The information of this clone is available through the following
URL.
http://xenopus.nibb.ac.jp.
Location/Qualifiers
1..616
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="XLO90j06"
/tissue_type="whole embryo"
/dev_stage="stage 25"
/clone_1id="NIBB Mochii normalized Xenopus tailbud
library"

ORIGIN

Query Match 47.3% Score 614.4; DB 4; Length 616;
Best local similarity 99.8%; Pred. No. 1.3e-181;
Matches 615; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Oy 264 GGTAGCCGACGATTCGTGCGAAGAGGCTGCCGACGTCGTGTCAGGTCGGATTAT 323
Db 616 GGTAGCCGACGATTCGTGCGAAGAGGCTGCCGACGTCGTGTCAGGTCGGATTAT 557

Query Match	44.7%	Score 580.8	DB 9	Length 853
Best Local Similarity	98.8%	Pred. No.5,4e-171		
Matches 585	Conservative 0	Mismatches 7	Indels 0	Gaps 0
QY	708	ATCAATGCTGACGGAGATATTTCTCTCTGGAACAAGACAGGGAATCCGGAGCCGGGGTG	767	
DB	1	ATCAATGCTGACGGAGATATTTCTCTCTGGAACAAGACAGGGAATCCGGAGCCGGGGTG	60	
QY	768	GGGAAGGATACCGGATTTCAACACAGTGGAAACCTTGTCAATGTTTGCATTAACCGGCAATT	827	
DB	61	GGGAAGGATACCGGATTTCAACACAGTGGAAACCTTGTCAATGTTTGCATTAACCGGCAATT	120	
QY	828	TTATTTGCTPACAAACGACGCGCAGAGGTTGGCCGACGCGCCGACCCCGTTATTAAGATT	887	
DB	121	TTATTTGCTPACAAACGACGCGCAGAGGTTGGCCGACGCGCCGACCCCGTTATTAAGATT	190	
QY	888	GATCAGACAGCGGTTTACCGCCCATCAACGCGCAAAAACAGCGGTATGTTGACATTACC	947	
DB	181	GATCAGACAGCGGTTTACCGCCCATCAACGCGCAAAAACAGCGGTATGTTGACATTACC	240	
QY	948	CACCTTCAAGCTGTTTATGCGCGGACGACGATTAATTCTGGCAAACTCTGGCGGCGCACT	1007	
DB	241	CACCTTCAAGCTGTTTATGCGCGGACGACGATTAATTCTGGCAAACTCTGGCGGCGCACT	300	
QY	1008	GGAGCTTCAACTGACGCGCTTCCGGTACGCGGATTAACAGCGCCGACAGTGTGTAACCTGT	1067	
DB	301	GGAGCTTCAACTGACGCGCTTCCGGTACGCGGATTAACAGCGCCGACAGTGTGTAACCTGT	360	
QY	1068	GTTTGAACGCTGGCGCTGCGGTACGATTAACGACGAGTGTGTAACGAGTTTCAGGTTTCGCTGCTT	1127	
DB	361	GTTTGAACGCTGGCGCTGCGGTACGATTAACGACGAGTGTGTAACGAGTTTCAGGTTTCGCTGCTT	420	
QY	1128	CCAGAATTTCACAGCATGCGGTATTAACGCGCGTGTCAATTAAATACGCGCGCCGGAGA	1187	
DB	421	CCAGAATTTCACAGCATGCGGTATTAACGCGCGTGTCAATTAAATACGCGCGCCGGAGA	480	
QY	1188	GGTGAAACTGACCCCTGGCAGAGATGTGAAGACGAAATCGCCAGGGCATGTCTTCTGTGGC	1247	
DB	481	GGTGAAACTGACCCCTGGCAGAGATGTGAAGACGAAATCGCCAGGGCATGTCTTCTGTGGC	540	
QY	1248	AGGTTTTCAGCGCAATCGTGAATGAAGACGCAATACCGCGCTGTGAGTTGTGTA	1299	
DB	541	AGGTTTTCAGCGCAATCGTGAATGAAGACGCAATACCGCGCTGTGAGTTGTGTA	592	
RESULT 3	CA093060	746 bp	mRNA	linear
LOCUS	CA093060			EST 23-SEP-2003
DEFINITION	SCCCCL2001D10.b Cl12 Saccharum officinarum cDNA clone SCCCL2001D10			
ACCESSION	CA093060			
VERSION	CA093060.1	GI:34946367		
KEYWORDS	EST.			
SOURCE	Saccharum officinarum			
ORGANISM	Saccharum officinarum			
REFERENCE	1 (bases 1 to 746)			
AUTHORS	Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.			
JOURNAL	The libraries that made SUCSEST			
COMMENT	Genet. Mol. Biol. 24 (1-4), 1-7 (2001)			
	Contact: Arruda P			

Centro de Biologia Molecular e Engenharia Genetica
Universidade Estadual de Campinas
Caixa Postal 6010, 13083-970, Campinas SP, Brazil
Tel: 55 19 3788 1137
Fax: 55 19 3788 1089
Email: pattud@unicamp.br
Clone distribution: clone distribution information can be found
through the Brazilian Clone Collection Center (BCCC) at
http://www.bcccenter.fcav.unesp.br
Plate: 001 row: D column: 10
Seq primer: M13/Forward primer.
Location/Qualifiers

FEATURES

source

1. /746
/organism="Saccharum officinarum"
/mol_type="mRNA"
/db_xref="taxon:4547"
/clone="SCCCL2001D10"
/lab_host="X11Blue MRP"
/clone_1lb="CL2"
/note="Organ: Pool of sugarcane calli submitted to low
(40C) and high (37 C) temperature stress; Vector:
pBluscript; Site 1: EcoRI; Site 2: XhoI; An
unidirectional cDNA library generated from [Pool of
sugarcane calli submitted to low (40C) and high (37 C)
temperature stress]. cDNA was prepared from polyA+ mRNA
using ZAP - cDNA Synthesis Kit (Stratagene). The
double-strand cDNAs were fractionated in a sepharose CL-2B
40cm-columns and fragments sizing between 0.8 and 1.5 Kb
were directionally cloned into the vector. Details of
each source of RNA and library construction can be
obtained at http://bucest.lad.fc.unicamp.br/public"

ORIGIN

Query Match 41.3%; Score 536.2; DB 6; Length 746;

Best Local Similarity 97.9%; Pred. NO. 5.8e-157;

Matches 552; Conservative 0; Mismatches 11; Indels 1; Gaps 1;

```

595 TTGTCCTTAAAGCTGAGAAACAGAGCAAGAGCTGTTTCATTAAAGCAGGATTACCATCG 654
184 TTGCGTACTGTCACTGAGAAACAGAGCAAGAGCTGTTTCATTAAAGCAGGATTACCATCG 243
655 GAACTCAAGGTGAGCGCCGACATGTTCTATTAAACCGGTGCGTTAAGCTTCGATCATG 714
244 GAACTCAAGGTGAGCGCCGACATGTTCTATTAAACCGGTGCGTTAAGCTTCGATCATG 303
715 CTGACGAGATATTTCTCTGCAACAGCAAGGAAATCCGGACCGGGGTGGGAGG 774
304 CTGACGAGATATTTCTCTGCAACAGCAAGGAAATCCGGACCGGGGTGGGAGG 363
775 ATCACCAGATTCACACAGTGAACACCTTGTGTAAGTTGCAATACGCGCAATTTTATTTG 834
364 ATCACCAGATTCACACAGTGAACACCTTGTGTAAGTTGCAATACGCGCAATTTTATTTG 423
835 CTACACGACGACGACGAGGTTCCCGACGCCGCCACCCCGTTATTAGATTGATCAG 894
424 CTACACGACGACGACGAGGTTCCCGACGCCGCCACCCCGTTATTAGATTGATCAG 483
895 ACAGCGTTGAGACCCCATCCACCCCAAAAAGAGCGTATGATGACATTACCACTTCA 954
484 ACAGCGTTGAGACCCCATCCACCCCAAAAAGAGCGTATGATGACATTACCACTTCA 543
955 GTGCTGTTTATCGCGACGACGATATCTGCAATCTCGCGCGCGCACTGGAGCTC 1014
544 GTGCTGTTTATCGCGACGACGATATCTGCAATCTCGCGCGCGCACTGGAGCTC 603
1015 AACTGAGCGCTTCCCGGTACCGCGATTAACGCGCGCAAGGTGTGAATCTGCTTTGAA 1074
604 AACTGAGCGCTTCCCGGTACCGCGATTAACGCGCGCAAGGTGTGAATCTGCTTTGAA 662
1075 CGCTGGCGCTCGGATTAAGGATTAACGCGCGATTAAGGATTAAGGATTAAGGATTAAG 1134
663 CGCTGGCGCTCGGATTAAGGATTAACGCGCGATTAAGGATTAAGGATTAAGGATTAAG 722

```

QY 1135 TTACAGCAGATCGTGATTAACG 1158
DB 723 TTACAGCANATCGTGATTAACG 746

RESULT 4

BU713770

LOCUS DEFINITION

BU713770 529 bp mRNA linear EST 23-OCT-2003
SUAHB02 Adult SJC 7/94 Schistosoma japonicum cDNA similar to
pddlIDK1A Chain A, Crystal Structure Of Phytate Complex Of
Escherichia Coli Phytase At Ph 6.6. Phytate Is Bound With Its
3-Phosphate In The Active Site. H92+ Cation Acts As An
Intermolecular Bridge, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PubMed

COMMENT

Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919 (ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn.

FEATURES

source

1. 529
/organism="Schistosoma japonicum"
/mol_type="mRNA"
/strain="Chinese (Anhui) strain"
/db_xref="taxon:6182"
/sex="Male and female"
/tissue_type="Whole body"
/dev_stage="Adult worms"
/lab_host="Mouse and rabbit"
/clone_1lb="Adult SJC 7/94"
/note="Vector: Lambda ZAP-II XR; Site 1: EcoRI; Site 2:
XhoI I; Several hundred adult Schistosoma japonicum
(Anhui, P.R. China, strain), of mixed sex, were perfused
from the mesenteries of experimentally infected mice and
rabbits at the Queensland Institute of Medical Research,
Brisbane, Australia (QIMR), and stored for several months
in liquid nitrogen. Subsequently, mRNA was isolated at the
QIMR from lysates of these worms by oligo dt
chromatography, using a kit from Pharmacia. The mRNA was
then shipped to Clontech, Palo Alto, CA, USA, who
constructed a cDNA library. First strand synthesis was
primed with an oligo-dt-XhoI primer and synthesized using
M-MuV reverse transcriptase. Second strand synthesis was
accomplished with RNase H and T4 DNA polymerase. The
double stranded cDNA was ligated to EcoRI linkers,
digested with EcoRI and XhoI, and ligated into the
phagemid vector lambda ZAP II XR. After construction of
this directional library by Clontech, it was returned to
the QIMR. During analysis of the library at the QIMR, we
have found that a small percentage, 2% to 3%, of the
clones contain inserts that appear to be highly homologous
to sequences from salmonid fishes, as determined by
homology comparisons using BLAST and by Southern
hybridization analysis to genomic DNA from salmon (Sigma
Chemical Co., St. Louis, MO) under stringent washing

conditions. The remainder of the clones appear to contain
S. japonicum sequences."

Query Match 40.7%; Score 529; DB 5; Length 529;
Best Local Similarity 100.0%; Pred. No. 9.8e-155;
Matches 529; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ORIGIN

670 GCCGACAAATGCTCTATTACCGGTGGGTAAAGCTTGGATCAATGTCAGCGAGATATTT 729
Db 1 GCCGAAATAGTCTCATTAAACGGTGGGTAAAGCTTGGATCAATGTCAGCGAGATATTT 60
Qy 730 CTCCTGCAACAAGCAGAGGAATGCGGAGCGGGGTGGGAAGATCAACGATTCACAC 789
Db 61 CTCCTGCAACAAGCAGAGGAATGCGGAGCGGGGTGGGAAGATCAACGATTCACAC 120
Qy 790 CAGTGGAACACTTGTCTTAAGTTGCAATACGGCAATTTTATTTGCTACAGCAGCCCA 849
Db 121 CAGTGGAACACTTGTCTTAAGTTGCAATACGGCAATTTTATTTGCTACAGCAGCCCA 180
Qy 850 GAGGTGCCCCGAGCGCGCCACCCGCTATTAGATTATATCAAGCAGCGTTAGCGCC 909
Db 181 GAGGTGCCCCGAGCGCGCCACCCGCTATTAGATTATATCAAGCAGCGTTAGCGCC 240
Qy 910 CATCCACCGCAAAAAGAGCGTATGCTGACATTACCACTTCACTGCTTTATCGCC 969
Db 241 CATCCACCGCAAAAAGAGCGTATGCTGACATTACCACTTCACTGCTTTATCGCC 300
Qy 970 GGAACGATTAATCTGCAATCTGCGCGCGGCACTGAGCTCACTGACGCTTCCC 1029
Db 301 GGAACGATTAATCTGCAATCTGCGCGCGGCACTGAGCTCACTGACGCTTCCC 360
Qy 1030 GGTGACCGCGATTAACACCGCGCAGGTGTAACCTGCTTTGAAGCGTGGCGTCCGTA 1089
Db 361 GGTGACCGCGATTAACACCGCGCAGGTGTAACCTGCTTTGAAGCGTGGCGTCCGTA 420
Qy 1090 AGCGATTAACAGCAGGATTCAGTTGCTGCTGCTTCCAGACTTACAGAGATGCT 1149
Db 421 AGCGATTAACAGCAGGATTCAGTTGCTGCTGCTTCCAGACTTACAGAGATGCT 480
Qy 1150 GATTAACAGCAGGATTCAGTTGCTGCTGCTTCCAGACTTACAGAGATGCT 1198
Db 481 GATTAACAGCAGGATTCAGTTGCTGCTGCTTCCAGACTTACAGAGATGCT 529

RESULT 5
CF326092/c 595 bp mRNA linear EST 18-AUG-2003
LOCUS JMT1--05-B09-g1 AUMT-overexpressing transgenic rice lambda phase
DEFINITION cDNA library (JMT1) Oryza sativa (japonica cultivar-group) cDNA
clone JMT1--05-B09, mRNA sequence.

ACCESSION CF326092
VERSION CF326092.1 GI:33800445
KEYWORDS EST.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1 (bases 1 to 595)
AUTHORS Kim,J.S., Jun,K.M., Cheong,P.J., Kim,M.J., Lee,T.H., Shin,Y.C.,
Song,S.I., Kim,J.K., Kim,Y.-K. and Nahm,B.H.
TITLE Large-scale Sequencing Analysis of Rice ESTs
JOURNAL Unpublished (2003)
COMMENT Contact: Nahm B.H.
Genomics and Genetics Institute, GreenGene Biotech Inc.; Division
of Bioscience and Bioinformatics, Myongji University
Yongin, Kyeonggi, Korea
Tel: 82 31 330 6193
Fax: 82 31 321 6355
Email: bhnahm@gbio.com, bhnahm@bio.myongji.ac.kr.

FEATURES
Location/Qualifiers
1..595

/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cui_fvar="Nackdong"
/db_xref="taxon:39947"
/clone="JMT1--05-B09"
/tissue_type="leaf"
/dev_stage="14 days after germination"
/lab_host="E.coli SDR"
/clone_id="AUMT-overexpressing transgenic rice lambda
phase cDNA library (JMT1)"
/note="Vector: pBluescript SK(+), Site 1, EcoRI, Site 2,
XhoI; cDNA was inserted into lambda Uni-ZAP XR vector at 5'
end with EcoRI and 3' end with XhoI site. mRNA was
prepared from Arabidopsis Jasmonate Carboxyl
methyltransferase overexpression line."

ORIGIN

Query Match 37.6%; Score 488.8; DB 7; Length 595;
Best Local Similarity 99.6%; Pred. No. 4.6e-142;
Matches 490; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

808 AGTTGCAATAGCGGCATTTTATTTGCTACAGCAGCAGAGTGGCCGCGCGCC 867
Db 595 AGTTGCAATAGCGGCATTTTATTTGCTACAGCAGCAGAGTGGCCGCGCGCC 536
Qy 868 GCCACCCCGTTATTAAGTTGATCAAGCAGCGTTGAGCCCATCCACCGCAAAAAG 927
Db 535 GCCACCCCGTTATTAAGTTGATCAAGCAGCGTTGAGCCCATCCACCGCAAAAAG 476
Qy 938 GCGATGATGATGATTAATCCCACTTCACTGCTGTTATGCGCGGACAGATTAATCTG 987
Db 475 GCGATGATGATGATTAATCCCACTTCACTGCTGTTATGCGCGGACAGATTAATCTG 416
Qy 988 GGAATTCCTGCGCGGCACTGAGCTCACTGAGAGCTTCCCGGTCACCGGATTAACG 1047
Db 415 GGAATTCCTGCGCGGCACTGAGCTCACTGAGAGCTTCCCGGTCACCGGATTAACG 356
Qy 1048 CCGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1107
Db 355 CCGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 296
Qy 1108 ATTCAAGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1167
Db 295 ATTCAAGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 236
Qy 1168 TTAATATAGCGCGCCCGGAGAGGTGAATGAACTGACCTTGGAGATGGAAGGAATCG 1227
Db 235 TTAATATAGCGCGCCCGGAGAGGTGAATGAACTGACCTTGGAGATGGAAGGAATCG 176
Qy 1228 CAGGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1287
Db 175 CAGGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 116
Qy 1288 TGCAGTTGTAA 1299
Db 115 TGCAGTTGTAA 104

RESULT 6
CN762997/c 895 bp mRNA linear EST 20-MAY-2004
LOCUS ID0AAASDC09BRI AgMS Acyrthosiphon pisum cDNA clone ID0AAASDC09 5',
DEFINITION mRNA sequence.

ACCESSION CN762997
VERSION CN762997.1 GI:47536920
KEYWORDS EST.
SOURCE Acyrthosiphon pisum (pea aphid)
ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha; Aphidiformes;
Aphidoidea; Aphididae; Macrosiphini; Acyrthosiphon.

REFERENCE 1 (bases 1 to 895)
AUTHORS Hunter,W., Martinez-Torres,D., Rahbe,Y., Sabater-Munoz,B.,

TITLE Stern, D., Tagu, D. and Wincker, P.
An expressed sequence tags database for the pea aphid *Acyrthosiphon pisum*
JOURNAL Unpublished (2004)
COMMENT Contact: D. Tagu
INRA Rennes
UMR BIO3P, BP 35327, F-35653 Le Rheu Cedex France
Tel: +33.2.23.48.51.65
Fax: +33.2.23.48.51.50
Risk of contamination by bacterial sequences from obligatory (Buchnera) or facultative endosymbionts. These sequences were obtained in the frame of the International Consortium of Aphid Genomics in collaboration with Genoscope

PCR Primers
FORWARD: CAGGAAACAGCTATGACC
Plate: 5
Location/Qualifiers
1. .895
/organism="Acyrthosiphon pisum"
/mol_type="mRNA"
/cultivar="developmentstage"
/db_xref="taxon:7029"
/clone="ID0AAA5DC09"
/tissue_type="whole insect"
/dev_stage="nymphs and adults (parthenogenetic females)"
/lab_host="XLI-Blue"
/note="Vector: pBS-SK minus; Site 1: EcoRI; Site 2: XhoI;
Sample name: ID0AAA ; Plant growth place: Department of Ecology & Evolutionary Biology, Princeton University ;
Soil conditions: Soil ; Sowing date: 01/06/1999 ;
Harvesting date: 01/06/1999 ; Stress date: no stress ;
Description: Aphids inoculated on one-week old *Vicia faba* under non-sterile conditions. All parthenogenetic stages and both winged and wingless adults were collected for library construction. ; experimental condition: long photoperiod (16-hr light/8-hr dark at 18 °C)"

ORIGIN

Query Match 37.2%; Score 483.8; DB 7; Length 895;
Best Local Similarity 99.6%; Pred. No. 1.9e-140;
Matches 485; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

813 GCATACGCGCAATTTATTTGCTACACGACCGCAGAGTTGCCGCGACCGCCGAC 872
895 GCATACGCGCAATTTATTTGCTACACGACCGCAGAGTTGCCGCGACCGCCGAC 836
873 CCCGTTTATTTGATTTGATCAAGACAGCGTTGACCGCCCATGACCGCAAAACAGGGGTA 932
835 CCCGTTTATTTGATTTGATCAAGACAGCGTTGACCGCCCATGACCGCAAAACAGGGGTA 776
933 TGGTGTACATTACCACTTCAAGTCTGTTTATGCGCGGACAGATATTAATCTGGCAAA 992
775 TGGTGTACATTACCACTTCAAGTCTGTTTATGCGCGGACAGATATTAATCTGGCAAA 716
993 TCTGCGGCGGCGCACTGAGCTCAACTGAGCCCTTCCGCGTACGCGGATTAACAGCGCGCC 1052
715 TCTGCGGCGGCGCACTGAGCTCAACTGAGCCCTTCCGCGTACGCGGATTAACAGCGCGCC 656
1053 AGGTGTGAACTGTGTTTGAACGCTGCGCTGCGTAAAGGATTAACAGCGCGGATTTCA 1112
655 AGGTGTGAACTGTGTTTGAACGCTGCGCTGCGTAAAGGATTAACAGCGCGGATTTCA 596
1113 GGTTCGCTGCTCTTCAGACTTTTACAGCAGATGCGTATTAACAGCGCGCTGTGATTTAA 1172
595 GGTTCGCTGCTCTTCAGACTTTTACAGCAGATGCGTATTAACAGCGCGCTGTGATTTAA 536
1173 TACGCGCGCGGAGAGTGAATCTGACCTTGCGCAGATGTAAGAGCAATTCGCGAGG 1232
535 TACGCGCGCGGAGAGTGAATCTGACCTTGCGCAGATGTAAGAGCAATTCGCGAGG 476
1233 CATGTGCTGTTGGAGGTTTAAAGCAATGCTGAATTAAGACGATACCGCGTTGAG 1292

Db 475 CATGTGCTGTTGGAGGTTTAAAGCAATGCTGAATGAAGACGATACCGCGCTGAG 416
Qy 1293 TTTGTAA 1299
Db 415 TTTGTAA 409

RESULT 7
CN754382/c 868 bp mRNA linear EST 19-MAY-2004
LOCUS ID0AAA12DE01RM1 ApMS Acyrthosiphon pisum cdna clone ID0AAA12DE01
DEFINITION 5', mRNA sequence.
ACCESSION CN754382
VERSION CN754382.1 GI:47519379
KEYWORDS EST.
SOURCE Acyrthosiphon pisum (pea aphid)
ORGANISM Acyrthosiphon pisum
Acyrthosiphon pisum
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha; Aphidiformes;
Aphidoidea; Aphididae; Macrosiphini; Acyrthosiphon.

REFERENCE 1 (bases 1 to 868)
Hunter, W., Martinez-Torres, D., Raabe, Y., Sabater-Munoz, B., Stern, D., Tagu, D. and Wincker, P.
An expressed sequence tags database for the pea aphid *Acyrthosiphon pisum*
JOURNAL Unpublished (2004)
COMMENT Contact: D. Tagu
INRA Rennes
UMR BIO3P, BP 35327, F-35653 Le Rheu Cedex France
Tel: +33.2.23.48.51.65
Fax: +33.2.23.48.51.50
Risk of contamination by bacterial sequences from obligatory (Buchnera) or facultative endosymbionts. These sequences were obtained in the frame of the International Consortium of Aphid Genomics in collaboration with Genoscope

PCR Primers
FORWARD: CAGGAAACAGCTATGACC
Plate: 12
row: E
column: 1.
Location/Qualifiers
1. .868
/organism="Acyrthosiphon pisum"
/mol_type="mRNA"
/cultivar="developmentstage"
/db_xref="taxon:7029"
/clone="ID0AAA12DE01"
/tissue_type="whole insect"
/dev_stage="nymphs and adults (parthenogenetic females)"
/lab_host="XLI-Blue"
/note="Vector: pBS-SK minus; Site 1: EcoRI; Site 2: XhoI;
Sample name: ID0AAA ; Plant growth place: Department of Ecology & Evolutionary Biology, Princeton University ;
Soil conditions: Soil ; Sowing date: 01/06/1999 ;
Harvesting date: 01/06/1999 ; Stress date: no stress ;
Description: Aphids inoculated on one-week old *Vicia faba* under non-sterile conditions. All parthenogenetic stages and both winged and wingless adults were collected for library construction. ; experimental condition: long photoperiod (16-hr light/8-hr dark at 18 °C)"

FEATURES

source

Query Match 34.3%; Score 445.8; DB 7; Length 868;
Best Local Similarity 99.3%; Pred. No. 1.8e-128;
Matches 458; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

839 AACGACGCGCAGAGTTGCCGCGCAGCGCCCAACCGTTATTTAGATTGATCAAGACAG 898
868 AACGACGCGCAGAGTTGCCGCGCAGCGCCCAACCGTTATTTAGATTGATCAAGACAG 810
899 CGTTAGCGCCCATTCACCGCAAAACAGCGGTTGTTGATCACTTACCACTTCAAGTC 958
809 CGTTAGCGCCCATTCACCGCAAAACAGCGGTTGTTGATCACTTACCACTTCAAGTC 750

QY 959 TGTATTGCGCGACACGATCTAATCTGGCAATCTCGCGCGCGCACTGAGCTCAACT 1018
 749 TGTATTGCGCGACACGATCTAATCTGGCAATCTCGCGCGCGCACTGAGCTCAACT 690
 QY 1019 GGAAGCTTCGCGTACGCGGATTAACAGCCGCCGAGGTGTAACTGGTGTGAAGCCT 1078
 689 GGAAGCTTCGCGTACGCGGATTAACAGCCGCCGAGGTGTAACTGGTGTGAAGCCT 630
 QY 1079 GGGGTGGGTAAAGCGATTAACAGCGAGTTCAGGTTTCGCTGGTCTTCCAGCTTAC 1138
 629 GGGGTGGGTAAAGCGATTAACAGCGAGTTCAGGTTTCGCTGGTCTTCCAGCTTAC 570
 QY 1139 AGCAGATGCGTGAATAAAGCGCGTGTCAATTAATACCGCCCGGAGAGGTGAACCTGA 1198
 569 AGCAGATGCGTGAATAAAGCGCGTGTCAATTAATACCGCCCGGAGAGGTGAACCTGA 510
 QY 1199 CCCTGGCAGATGTGAAGAGCGCAATGCGGAGGCGATGTTTCGTTGGCAGGTTTAAACG 1258
 509 CCCTGGCAGATGTGAAGAGCGCAATGCGGAGGCGATGTTTCGTTGGCAGGTTTAAACG 450
 QY 1259 AATTCGTGAATGAAGACGCGATACCGCGCTTGACGTTTGA 1299
 449 AATTCGTGAATGAAGACGCGATACCGCGCTTGACGTTTGA 409
 Db

RESULT 8
 CN757565/c 864 bp mRNA linear EST 20-MAY-2004
 LOCUS ID0AA1D09RM1 ApMS Acyrthosiphon pisum cDNA clone ID0AA1D09 5',
 DEFINITION mRNA sequence.
 ACCESSION CN757565
 VERSION CN757565.1 GI:47531488
 KEYWORDS EST.
 SOURCE Acyrthosiphon pisum (pea aphid)
 ORGANISM Acyrthosiphon pisum
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha; Aphidiformes;
 Aphidoidea; Aphididae; Macrosiphini; Acyrthosiphon.
 1 (bases 1 to 864)
 Hunter,W., Martinez-Torres,D., Rahbe,Y., Sabater-Munoz,B.,
 Stern,D., Tagu,D. and Winkler,P.
 An expressed sequence tags database for the pea aphid Acyrthosiphon

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 UNPUBLISHED (2004)
 Contact: D. Tagu
 INRA Rennes
 UMR BIO3P, BP 35327, F-35653 Le Rheu Cedex France
 Tel: +33.2.23.48.51.65
 Fax: +33.2.23.48.51.50

FEATURES
 source
 Plate: 1 row: F column: 9.
 Location/Qualifiers
 1..864
 /organism="Acyrthosiphon pisum"
 /mol_type="mRNA"
 /cultivar="developmentstage"
 /db_xref="taxon:7029"
 /clone="ID0AA1D09"
 /rname_type="whole insect"
 /dev_stage="nymphs and adults (parthenogenetic females)"
 /lab_host="X11-Blue"
 /clone_lib="ApMS"
 /note="Vector: pBS-SK minus; Site 1: EcoRI; Site 2: XhoI;
 Sample name: ID0AA1; Plant growth place: Department of
 Ecology & Evolutionary Biology, Princeton University ;
 Soil conditions: Soil ; Sowing date: 01/06/1999 ;
 Harvesting date: 01/06/1999 ; Stress date: no stress ;
 Description: Aphids inoculated on one-week old *Vicia faba*

ORIGIN

Query Match 34.0%; Score 441.8; DB 7; Length 864;
 Best Local Similarity 99.3%; Pred. No. 3,3e-127;
 Matches 454; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

under non-sterile conditions. All parthenogenetic stages
 and both winged and wingless adults were collected for
 library construction. ; experimental condition: long
 photoperiod (16-hr light/8-hr dark at 18 °C)"

QY 843 CAGCCAGAGTTGCCCGCAGCCCGCCACCCCGTATTAGATTGATCAAGACAGCTT 902
 864 CAGCCAGAGTTGCCCGCAGCCCGCCACCCCGTATTAGA-TTGAATCAAGACAGCTT 806
 QY 903 GAGCCCATCAACCGCAAAAGAGCGTATGTTGACATTAACCACTTCACTGCTTT 962
 805 GAGCCCATCAACCGCAAAAGAGCGTATGTTGACATTAACCACTTCACTGCTTT 746
 QY 963 TATCGCGGACACGATCTAATCTGCGCAATCTCGGCGGCGCATGAGCTCAACTGGAC 1022
 745 TATCGCGGACACGATCTAATCTGCGCAATCTCGGCGGCGCATGAGCTCAACTGGAC 686
 QY 1023 GCTTCCCGTACGCGGATTAACAGCCGCCAGGTGTGAATGTTTGAACGCTGGCG 1082
 685 GCTTCCCGTACGCGGATTAACAGCCGCCAGGTGTGAATGTTTGAACGCTGGCG 626
 QY 1083 TCGGCTAAGCATTAACAGCCAGTGAATGAGTTTCGCTGCTTCCAGCTTTACAGCA 1142
 625 TCGGCTAAGCATTAACAGCCAGTGAATGAGTTTCGCTGCTTCCAGCTTTACAGCA 566
 QY 1143 GATGCGTATTAACAGCCGCTGTCTAATTAAGCGCGCCGAGAGGTGAACAGCTT 1202
 565 GATGCGTATTAACAGCCGCTGTCTAATTAAGCGCGCCGAGAGGTGAACAGCTT 506
 Db 1203 GGCAGATGTGAAGAGCGCAATGCGCAGGCGATGTTTCGTTGACGTTTACGCAAT 1262
 505 GGCAGATGTGAAGAGCGCAATGCGCAGGCGATGTTTCGTTGACGTTTACGCAAT 446
 QY 1263 CGTGAATGAAGACGCGATACCGCGCTTGACGTTTGA 1299
 445 CGTGAATGAAGACGCGATACCGCGCTTGACGTTTGA 409
 Db

RESULT 9
 BE520240/c 393 bp mRNA linear EST 19-MAR-2001
 LOCUS M1B12STM Arabidopsis developing seed Arabidopsis thaliana cDNA
 DEFINITION clone M1B12 5', mRNA sequence.
 ACCESSION BE520240
 VERSION BE520240.1 GI:9778242
 KEYWORDS EST.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 393)
 White,J.A., Todd,J., Newman,T., Rocks,N., Girke,T., Martinez de
 Ilarduya,O., Jaworski,J.G., Ohlrogge,J. and Benning,C.
 A new set of Arabidopsis expressed sequence tags from developing
 seeds. The metabolic pathway from carbohydrrates to seed oil
 plant Physiol. 124 (4), 1582-1594 (2000)

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 MEDLINE
 PUBMED
 Contact: Benning, C
 Dept. of Biochemistry & Molecular Biology
 Michigan State University
 224 Biochemistry, Michigan State University, East Lansing, MI
 48824, USA
 Tel: 517 355 1609
 Fax: 517 353 9334
 Email: benning@msu.edu
 Michigan State University DNA Sequencing Facility Arabidopsis

Biological Resource Center, The Ohio State University, 309 Botany &
Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX:
6142920603 TEL: 6142929371.
Location/Qualifiers

FEATURES

source

1. .393
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/ecotype="Columbia"
/db_xref="taxon:3702"
/clone="M11812"
/issue_type="seed"
/dev_stage="5-13 days after flowering"
/lab_host="E.coli"
/clone_1lb="Arabidopsis developing seed"
/note="Organ: Developing seed; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoI"

ORIGIN

Query Match 29.7%; Score 386.2; DB 2; Length 393;
Best Local Similarity 98.7%; Pred. No. 8.6e-110;
Matches 388; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 319 ATTATGCTGATGTCGACGAGCGTACCCGTTAAACAGCGAGCGCTTCGCGCGGAGCTG 378
|||||
DB 393 ATTATGCTGATGTCGACGAGCGTACCCGTTAAACAGCGAGCGCTTCGCGCGGAGCTG 334
|||||
OY 379 GCACTGACTGTGCAATTAACCGTACATACCAGGAGATACGTCAGTCCGATCCGTTA 438
|||||
DB 333 GCACTGACTGTGCAATTAACCGTACATACCAGGAGATACGTCAGTCCGATCCGTTA 274
|||||
OY 439 TTTAATCTCTTAAACATGCGCTTTGCGCACTGATTAACCGAAGCGTACGCGATC 498
|||||
DB 273 TTTAATCTCTTAAACATGCGCTTTGCGCACTGATTAACCGAAGCGTACGCGATC 214
|||||
OY 499 CTCAGCAGGCGAGAGGGTCAATTCGATCTTAACGGGCGATCGGCAACGGCGTTTCG 558
|||||
DB 213 CTCAGCAGGCGAGAGGGTCAATTCGATCTTAACGGGCGATCGGCAACGGCGTTTCG 154
|||||
OY 559 GAACTGGAGCGGGTCTTAATTTTTCGCAATCAAACTTGCTTAAACGTGAGAAACAG 618
|||||
DB 153 GAACTGGAGCGGGTCTTAATTTTTCGCAATCAAACTTGCTTAAACGTGAGAAACAG 94
|||||
OY 619 GACGAAGAGCTGCTTAATTAACGAGGCTTACCATCGAATCAAGGTGAGCGCGCAAT 678
|||||
DB 93 GACGAAGAGCTGCTTAATTAACGAGGCTTACCATCGAATCAAGGTGAGCGCGCAAT 34
|||||
OY 679 GTCCTAATACCGGTGCGGTAAAGCTCGCATCA 711
|||||
DB 33 GTCCTAATACCGGTGCGGTAAAGCTCGCATCA 1
|||||

RESULT 10 383 bp mRNA linear EST 18-MAY-2001
LOCUS AM036132

DEFINITION EST274508 tomato seed, TAMU Lycopersicon esculentum cDNA clone
cLBE1E23 similar to periplasmic phosphoanhydride phosphohydrolase
precursor, putative, mRNA sequence.

ACCESSION AM036132
VERSION AM036132.1 GI:5894811

KEYWORDS

EST.

SOURCE

Lycopersicon esculentum (tomato)

ORGANISM

Lycopersicon esculentum

REFERENCE

AUTHORS

Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T.,
Holt,I.E., Liang,F., Upton,J., Ronning,C.M., Craven,M.B.,
Pujil,C.Y., Bowman,C.L., Nierman,W., Fraser,C.M., Venter,J.C.,
Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato seed tissue
Unpublished (1999)
Contact: CUGI

Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
3 prime sequence.
Location/Qualifiers

FEATURES

source

1. .383
/organism="Lycopersicon esculentum"
/mol_type="mRNA"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLBE1E23"
/issue_type="seeds"
/dev_stage="quiescent seed"
/lab_host="XLI-Blue MFP"
/clone_1lb="tomato seed, TAMU"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; cLBE - Tomato Seed EST library. Directionally cloned
cDNAs inserted into pBluescript SK(-) at 5' end with
EcoRI and 3' end with XhoI site."

ORIGIN

Query Match 29.5%; Score 383; DB 2; Length 383;
Best Local Similarity 100.0%; Pred. No. 8.7e-109;
Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 348 TAAACAGGCGAAGCCTTGCGCGCGGCTGGACCTGACTGTGCAATTAACCTACATAC 407
|||||
DB 1 TAAACAGGCGAAGCCTTGCGCGCGGCTGGACCTGACTGTGCAATTAACCTACATAC 60
|||||
OY 408 CCAGGAGATAGCTCCAGTCCGATCCGATCTTATTAATCTCTTAAACATGCGCTTGCCA 467
|||||
DB 61 CCAGGAGATAGCTCCAGTCCGATCCGATCTTATTAATCTCTTAAACATGCGCTTGCCA 120
|||||
OY 468 ACTGATTAACGCGAAGCTGACGATCCGATCTGACAGGCGAGAGGTCATTTCTGA 527
|||||
DB 121 ACTGATTAACGCGAAGCTGACGATCCGATCTGACAGGCGAGAGGTCATTTCTGA 180
|||||
OY 528 CTTTACCGGCGATCGCAACCGCGCTTGCGCAACTGGAACGGGTGCTTAATTTCCGA 587
|||||
DB 181 CTTTACCGGCGATCGCAACCGCGCTTGCGCAACTGGAACGGGTGCTTAATTTCCGA 240
|||||
OY 588 ATCAACTGTGCTTAAACGTGAGAAACGAGCAAGCTGTTCAATTAACGAGGACTT 647
|||||
DB 241 ATCAACTGTGCTTAAACGTGAGAAACGAGCAAGCTGTTCAATTAACGAGGACTT 300
|||||
OY 648 ACCATGGAATCAAGGTGAGCGCGCAATGTCTGATTAACCGGTGCGTAAAGCTCGC 707
|||||
DB 301 ACCATGGAATCAAGGTGAGCGCGCAATGTCTGATTAACCGGTGCGTAAAGCTCGC 360
|||||
OY 708 ATCAATGCTGACGAGATATTTC 730
|||||
DB 361 ATCAATGCTGACGAGATATTTC 383
|||||

RESULT 11 383 bp mRNA linear EST 18-MAY-2001
LOCUS AM036134/c

DEFINITION EST274510 tomato seed, TAMU Lycopersicon esculentum cDNA clone
cLBE1E23 similar to periplasmic phosphoanhydride phosphohydrolase
precursor, putative, mRNA sequence.

ACCESSION AM036134
VERSION AM036134.1 GI:5894813

KEYWORDS

EST.

SOURCE

Lycopersicon esculentum (tomato)

ORGANISM

Lycopersicon esculentum

REFERENCE

AUTHORS

Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T.,
Holt,I.E., Liang,F., Upton,J., Ronning,C.M., Craven,M.B.,
Pujil,C.Y., Bowman,C.L., Nierman,W., Fraser,C.M., Venter,J.C.,
Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato seed tissue
Unpublished (1999)
Contact: CUGI

TITLE Martin, G.B., Tanksley, S.D. and Giovannoni, J.
JOURNAL Generation of ESTs from tomato seed tissue
COMMENT Unpublished (1999)
Contact: CUGI
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
5 prime sequence.

FEATURES
source
1. 383
Location/Qualifiers

/organism="Lycopersicon esculentum"
/mol_type="mRNA"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="C1EB1E23"
/tissue_type="seeds"
/dev_stage="quiescent seed"
/lab_host="X11-Blue MPF"
/clone_lib="Tomato seed, TAMU"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; c1EB - Tomato Seed EST Library. Directionally cloned cDNAs inserted into pBluescript SK(-) at 5' end with EcoRI and 3' end with XhoI site."

ORIGIN

Query Match 29.5%; Score 383; DB 2; Length 383;
Best Local Similarity 100.0%; Pred. No. 8.7e-109;
Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 348 TAAAAAGGAGGAGCCTTCGCGCGCGGCTGACCTGACCTGACATTAACGTATAC 407
DB 383 TAAAAAGGAGGAGCCTTCGCGCGCGGCTGACCTGACCTGACATTAACGTATAC 324
QY 408 CCAGGAGATACGTCGACGATCCGATCCGATTTAATCTCTAATAAATGCGCTTGCA 467
DB 323 CCAGGAGATACGTCGACGATCCGATCCGATTTAATCTCTAATAAATGCGCTTGCA 264
QY 468 ACTGATTAACGGAACGTCGACGATCCGATCCGATTTAATCTCTAATAAATGCGCTTGCA 527
DB 263 ACTGATTAACGGAACGTCGACGATCCGATCCGATTTAATCTCTAATAAATGCGCTTGCA 204
QY 528 CTTACCGGCGCATCGGAAACGGGCTTGGCAACTGGAAGGGGTCAATTTTCGCA 587
DB 203 CTTACCGGCGCATCGGAAACGGGCTTGGCAACTGGAAGGGGTCAATTTTCGCA 144
QY 588 ATCAAACTTGCTTAAAGCTGAAGAAACGAAAGCTGTTCAATTAACGAGGAT 647
DB 143 ATCAAACTTGCTTAAAGCTGAAGAAACGAAAGCTGTTCAATTAACGAGGAT 84
QY 648 ACCATCGGAATCAAGGTGAGCGCGCAATGTCATTTAACCGGTGCGGTAGCCTCGC 707
DB 83 ACCATCGGAATCAAGGTGAGCGCGCAATGTCATTTAACCGGTGCGGTAGCCTCGC 24
QY 708 ATCAATCTGACGAGATATTTTC 730
DB 23 ATCAATCTGACGAGATATTTTC 1

RESULT 12
LOCUS BUT13769 531 bp mRNA linear EST 23-OCT-2003
DEFINITION SJAABUG01 Adult SJC 7/94 Schistosoma japonicum cDNA similar to
ep1p07102|PFA ECOLI PERIPLASMIC APPA PROTEIN PRECURSOR (INCLUDES:
PHOSPHOANHYDRIDE PHOSPHOHYDROLASE (PH 2.5 ACID PHOSPHATASE) (AP);
6-PHYTASE), mRNA sequence.

ACCESSION BUT13769
VERSION BUT13769.1 GI:28321125
KEYWORDS EST;
SOURCE Schistosoma japonicum
ORGANISM Schistosoma japonicum
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
Strigoidida; Schistosomatidae; Schistosomatidae; Schistosoma.

REFERENCE

1 (bases 1 to 531)
AUTHORS Hu, W., Yan, O., Shen, D. K., Liu, F., Zhu, Z. D., Song, H. D., Xu, X. R.,
Wang, Z. J., Rong, Y. P., Zeng, L. C., Wu, J., Zhang, X., Wang, J. J.,
Xu, X. N., Wang, S. Y., Fu, G., Zhang, X. L., Wang, Z. Q., Brindley, P. J.,
McManus, D. P., Xue, C. L., Feng, Z., Chen, Z. and Han, Z. G.
Evolutionary and biomedical implications of a Schistosoma japonicum
complementary DNA resource
Nat. Genet. 35 (2), 139-147 (2003)

JOURNAL MEDLINE PUBMED

12973349
Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919 (ex. 45)
Fax: 86-21-50801922
Email: hanzg@hgc.sh.cn

FEATURES

source

Location/Qualifiers

1. 531
/organism="Schistosoma japonicum"
/mol_type="mRNA"
/strain="Chinese (Aohui) strain"
/db_xref="taxon:6182"
/sex="Male and female"
/tissue_type="whole body"
/dev_stage="Adult worms"
/lab_host="Mouse and rabbit"
/clone_lib="Adult SJC 7/94"
/note="Vector: Lambda ZAP-II XR.; Site_1: EcoR I; Site_2: XhoI I; Several hundred adult Schistosoma japonicum (Aohui, P. R. China, strain), of mixed sex, were perfused from the mesenteries of experimentally infected mice and rabbits at the Queensland Institute of Medical Research, Brisbane, Australia (QIMR), and stored for several months in liquid nitrogen. Subsequently, mRNA was isolated at the QIMR from lysates of these worms by oligo dt chromatography, using a kit from Pharmacia. The mRNA was then shipped to Clontech, Palo Alto, CA, USA, who constructed a cDNA library. First strand synthesis was primed with an oligo-dT-XhoI-primer and synthesized using M-MuV reverse transcriptase. Second strand synthesis was accomplished with RNase H and T4 DNA polymerase. The double stranded cDNA was ligated to EcoRI linkers, digested with EcoRI and XhoI, and ligated into the phagemid vector lambda ZAP II XR. After construction of this directional library by Clontech, it was returned to the QIMR. During analysis of the library at the QIMR, we have found that a small percentage, 28 to 3%, of the clones contain inserts that appear to be highly homologous to sequences from salmonid fishes, as determined by homology comparisons using BLAST and by Southern hybridization analysis to genomic DNA from salmon (Sigma Chemical Co., St. Louis, MO) under stringent washing conditions. The remainder of the clones appear to contain S. japonicum sequences."

ORIGIN

Query Match 27.2%; Score 353.4; DB 5; Length 531;
Best Local Similarity 88.6%; Pred. No. 2e-99; 51; Indels 4; Gaps 4;
Matches 428; Conservative 0; Mismatches 51;

QY 670 GCCGCAATGCTCTAATTAACCGGTGGGTAGCCTCGCATCATGCTGACGAGATATTT 729
DB 1 GCCGCAATGCTCTAATTAACCGGTGGGTAGCCTCGCATCATGCTGACGAGATGCGG 60
QY 730 CTCCTGCAACAAGCAGGAAATGCCGAGCGGGGTGGGAAAGATCAACGATTACAC 789
DB 61 CTCCTGCAACAAGCAGGAAATGCCGAGCGGGGTGGGAAAGATCAACGATTACAC 120
QY 790 CAGTGAACACCTTGTTAGTTGCTTAACGGCAATTTATTTGCTACAGCAGCGCA 849
DB 121 CAGTGAACACCTTGTTAGTTGCTTAACGGCAATTTATTTGCTACAGCAGCGCA 180

Page 9

Db	351	TCCTTCAGACAGGGGACGGGA-GGCTCAATTGCTGTATTTACCGGGGATCTGGCAAAACGGGGCGTTC	29.9
OY	557	GCGAATCTGGAAACGGGTGCTTAATTTTCGCGAATCAAACTTGTCCTTAAACGTGAGAAAC	61.6
Db	292	GCGAACTGGAAACGGGTG-TTAAATTTTTCGCAATCAAACTGTGTGCTTTAAAGTGAAGAAC	23.4
OY	617	AGGACGAAAGCTGTTCAATTAAACGACGAGCATTAACCATCGAACTCAAGGTGAGCGCCGACA	67.6
Db	233	AGGACGAAAGCTGTTCAATTAAACGAGGACATTAACCATCGGAATCAAGTGAAGCCCGACA	17.4
OY	677	ATGTCCTATTAAACCGGTGCGGTAAAGCTTCGCATCAATGCTGACGGAATATTTCTCCTGC	73.6
Db	173	ATGTCCTATTAAACCGGTGCGGTAAAGCTTCGCATCAATGCTGACGGAATATTTTTCCTGC	11.4
OY	737	AACAAGACAAGGGAATGCCGGAAGCCGGGGGTGGGAAAGATACCGATTCAACACAGTGTGA	79.6
Db	113	AACAAGACAAGGGAATGCCGGAAGCCGGGGGTGGGAAAGATACACCGATTCAACACAGTGTGA	5.4
OY	797	ACACCTTGCTTAAGTTTGATTAACCCGCAATTTTATTTGTCTACAGAGCAGCGCA	84.9
Db	53	ACACCTTGCTTAAGTTTGATTAACCCGCAATTTTATTTGTCTACAGAGCAGCGCA	1

LOCUS	CN759004	706 bp	mRNA	linear	ESI 20-Fat-200
DEFINITION	ID0AA24BC04RM1 ApMs Acyrtosiphon pisum cDNA clone ID0AAA24BC04				
ACCESSION	CN759004				
VERSION	CN759004.1	GI:47532927			
KEYWORDS	5', mRNA sequence.				
FEATURES					

SOURCE	ORGANISM
ACYRTHOSIPHON pisum (pea aphid)	
ACYRTHOSIPHON pisum	
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha; Aphidiformes; Aphidoidea; Aphididae; Macroscaphini; ACYRTHOSIPHON.	

AUTHORS Hunter, M., Meléndez-Velazquez, D., Gaudin, L., Sabatier, H., Hanz, D., Stern, D., Tagu, D., and Wincker, P.

TITLE An expressed sequence tags database for the pea aphid *Acyrthosiphon pisum*

JOURNAL Unpublished (2004)

COMMENT Contact: D. Tagu

(Buchnera) or facultative endosymbionts. These sequences were obtained in the frame of the International Consortium of Aphid Genomics in collaboration with Genoscope
PCR primers
FORWARD: CAGGAACAGCTATGACC

```

/lab_host="XLI-Blue"
/clone_id="APMS"
/notes="Vector: pBS-SK minus. Site 1: EcoRI, Site 2: XhoI,
Sample name: ID04A, Plant growth place: Department of

```

Harvesting date: 01/06/1999 ; Stress date: no stress ; Description: Aphids inoculated on one-week old *Vicia faba* under non-sterile conditions. All parthenogenetic stages and both winged and wingless adults were collected for

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 21:17:58 ; Search time 6258.48 Seconds
(without alignments)
10057.285 Million cell updates/sec

Title: US-09-926-375B-7_COPY_12653_13951

Perfect score: 1299
Sequence: 1 atgaagaagcattcttaacc.....taccgcctgcagttgttaa 1299

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenBank1.*

1: gb_ha.*
2: gb_hcg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1299	100.0	3470	6	AX042376 Sequence
2	1299	100.0	4060	6	AX042374 Sequence
3	1299	100.0	6116	6	AX042375 Sequence
4	1299	100.0	6708	6	AX042373 Sequence
5	1299	100.0	17732	6	AX042378 Sequence
6	1299	100.0	20623	6	AX042372 Sequence
7	1294.2	99.6	18882	1	D90735 Escherichia
8	1294.2	99.6	110000	1	U00096_10 Continuation (11 o
9	1293.2	99.6	1901	6	ECOAPPA Escherichia
10	1293.2	99.6	1901	6	AX356572 Sequence
11	1291.6	99.4	1901	6	ECOACPBH Escherichia
12	1291.6	99.4	1901	6	AX356570 Sequence
13	1291.6	99.4	1901	6	AX356574 Sequence
14	1291	99.4	1315	1	AF537219 Escherich
15	1290	99.3	1901	1	ECOACPHC Escherichia
16	1290	99.3	1901	1	ECOACPHD Escherichia
17	1290	99.3	1901	1	ECOACPHE Escherichia
18	1290	99.3	1901	1	ECOACPHF Escherichia
19	1288.4	99.2	1901	1	ECOACPHA Escherichia

20	1288.4	99.2	1901	6	AX356571 Sequence
21	1286	99.0	5421	6	AX042377 Sequence
22	1282.2	98.7	1323	6	AR108133 Sequence
23	1282.2	98.7	1323	6	AR127818 Sequence
24	1282.2	98.7	1323	6	AR130956 Sequence
25	1282.2	98.7	1323	6	AR492966 Sequence
26	1282.2	98.7	1323	6	AX052713 Sequence
27	1282.2	98.7	1323	6	AX389666 Sequence
28	1282.2	98.7	1323	6	AX356566 Sequence
29	1267	97.5	11761	1	AE015127 Shigella
30	1265.4	97.4	292504	1	AE016981 Shigella
31	1257.4	96.8	1489	1	AY496073 Escherich
32	1257.4	96.8	1489	6	AR277753 Sequence
33	1246.2	95.9	10029	1	AE005292 Escherich
34	1246.2	95.9	327773	1	AP002554 Escherich
35	1237.4	95.3	301276	1	AE016758 Escherich
36	1170.2	90.1	1272	6	BD074747 Novel phy
37	716.6	55.2	212936	2	AC020970 Mus muscu
38	447.4	34.4	3279	2	AY390262 Citrobact
39	321.8	24.8	114979	2	AC020798 Mus muscu
40	305.2	23.5	760	1	AY378096 Obesumbac
41	303.8	23.4	2558	1	AY425982 Obesumbac
42	302	23.2	2161	1	AY425982 Obesumbac
43	228.6	17.6	10029	1	AE013783 Yersinia
44	228.6	17.6	110000	1	Continuation (29 o
45	228.6	17.6	193050	1	AJ414149 Yersinia

ALIGNMENTS

RESULT 1	AX042376	AY042376	3470 bp	DNA	linear	PAT 23-NOV-2000
LOCUS	AX042376	Sequence 5 from Patent WO0064247.				
DEFINITION	AX042376					
ACCESSION	AX042376					
VERSION	AX042376.1	GI:11340994				
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
1						
AUTHORS						
TITLE						
JOURNAL						
FEATURES						
source						
1..3470						
/organism="synthetic construct"						
/mol_type="unassigned DNA"						
/db_xref="taxon:32630"						
/note="R15/APPA transgene"						
ORIGIN						
Query Match						
Best Local Similarity						
Matches 1299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;						
QY	1	ATGAAGCCATCTTAATCCATTCTTTATCTCTGATTCGTTACCCGCAATCTGCA	60			
DB	1811	ATGAAGCCATCTTAATCCATTCTTTATCTCTGATTCGTTACCCGCAATCTGCA	1870			
QY	61	TTGCTCAGAGTGAGCCGAGCTGAGAGCTGGAAGTGTGTGATTCGTCATGCTGATGCT	120			
DB	1871	TTGCTCAGAGTGAGCCGAGCTGAGAGCTGGAAGTGTGTGATTCGTCATGCTGATGCT	1930			
QY	121	GTGGTGCTCCACCAAGGCGACGCACTGATGAGGATGTCACCCGACGATGAGCA	180			
DB	1931	GTGGTGCTCCACCAAGGCGACGCACTGATGAGGATGTCACCCGACGATGAGCA	1990			
QY	181	ACCTGGCCGATTAACCTGGTGGCTGACACCGCGCGTGGTGAAGCTTAATCGCTATCTC	240			
DB	1991	ACCTGGCCGATTAACCTGGTGGCTGACACCGCGCGTGGTGAAGCTTAATCGCTATCTC	2050			

QY 241 GACATTAACCAAGCCAGCGCTGTGGTAGCCGACGGAATTCCTGGCGAAAAAGGCGTCCG 300
DB 2051 GACATTAACCAAGCCAGCGCTGTGGTAGCCGACGGAATTCCTGGCGAAAAAGGCGTCCG 2110
QY 301 CAGTCTGGGCAAGTGGCGGATTTATGCTGATGTGACGAGCGTAAACCGGTAAAAACAGCGCA 360
DB 2111 CAGTCTGGGCAAGTGGCGGATTTATGCTGATGTGACGAGCGTAAACCGGTAAAAACAGCGCA 2170
QY 361 GCGTTGCGCCGCGGCTGGACCTGACCTGTGCAATTAACCGTACATTAACCGGAGCATAGC 420
DB 2171 GCGTTGCGCCGCGGCTGGACCTGACCTGTGCAATTAACCGTACATTAACCGGAGCATAGC 2230
QY 421 TCCAGTCCCGATCCGTTATTTATCTCTTAAAAAATGCGGCTTGGCACTGATTAACGG 480
DB 2231 TCCAGTCCCGATCCGTTATTTATCTCTTAAAAAATGCGGCTTGGCACTGATTAACGG 2290
QY 481 AACGTGACTGACGCGATCCTCAGCAGGAGGAGGATCAATTGCTGACTTTACCGGAGCAT 540
DB 2291 AACGTGACTGACGCGATCCTCAGCAGGAGGAGGATCAATTGCTGACTTTACCGGAGCAT 2350
QY 541 CGGCAAAACGCGCTTTCGCGAATGGAACGCGGTCTTAATTTTCGCAATCAACTGTGCG 600
DB 2351 CGGCAAAACGCGCTTTCGCGAATGGAACGCGGTCTTAATTTTCGCAATCAACTGTGCG 2410
QY 601 CTTAAACGTGAGAAACAGAGCAAGAGCTGTTCAATTAACGAGCATTAACCATCGAACTC 660
DB 2411 CTTAAACGTGAGAAACAGAGCAAGAGCTGTTCAATTAACGAGCATTAACCATCGAACTC 2470
QY 661 AAGGTGAGGCGCGACCAATGCTCATTAACCGGAGGCTTAAGCGCTGATCAATGCTGAGC 720
DB 2471 AAGGTGAGGCGCGACCAATGCTCATTAACCGGAGGCTTAAGCGCTGATCAATGCTGAGC 2530
QY 721 GAGATATTTCTCTGCAACAAGCAAGGAAATGCCGAGCCGCGGTGGGAAAGATCAC 780
DB 2531 GAGATATTTCTCTGCAACAAGCAAGGAAATGCCGAGCCGCGGTGGGAAAGATCAC 2590
QY 781 GATTACACACAGTGAACACCTTGCTAATGTTGCAATTAACCGCAATTTTATTTGCTACAA 840
DB 2591 GATTACACACAGTGAACACCTTGCTAATGTTGCAATTAACCGCAATTTTATTTGCTACAA 2650
QY 841 CGAGCGCCAGAGGTTGCGCGAGCGCGGCAACCGGTTATAGATTGATCAAGACAGCG 900
DB 2651 CGAGCGCCAGAGGTTGCGCGAGCGCGGCAACCGGTTATAGATTGATCAAGACAGCG 2710
QY 901 TTGACCGCCCATCAACCGCAAAAACAGCGTATGTTGATGACATTAACCACTTCACTGCTG 960
DB 2711 TTGACCGCCCATCAACCGCAAAAACAGCGTATGTTGATGACATTAACCACTTCACTGCTG 2770
QY 961 TTTATGCGCGGACAGATTAATCTGCGCAATCTGCGGCGGCACTGAGGCTCAACTG 1020
DB 2771 TTTATGCGCGGACAGATTAATCTGCGCAATCTGCGGCGGCACTGAGGCTCAACTG 2830
QY 1021 AGCGTTCCCGGTCAGCGCGGATTAACGCGCGCAGGTGTGAATGATGTTTGAACGCTG 1080
DB 2831 AGCGTTCCCGGTCAGCGCGGATTAACGCGCGCAGGTGTGAATGATGTTTGAACGCTG 2890
QY 1081 CGTGGCTAAGCGATTAACAGCCAGTGAATTCAGGTTTCGCTGCTTCCAGACTTTACAG 1140
DB 2891 CGTGGCTAAGCGATTAACAGCCAGTGAATTCAGGTTTCGCTGCTTCCAGACTTTACAG 2950
QY 1141 CAGATCGGTGATTAACCGCGCTGTCAATTAATACCGCGCCGAGAGGTTGAACGCTGAC 1200
DB 2951 CAGATCGGTGATTAACCGCGCTGTCAATTAATACCGCGCCGAGAGGTTGAACGCTGAC 3010
QY 1201 CTGGGAGAGTGAAGGGAATGCGAGGAGGATGTTGTTGAGGTTTACGAA 1260
DB 3011 CTGGGAGAGTGAAGGGAATGCGAGGAGGATGTTGTTGAGGTTTACGAA 3070
QY 1261 ATCGTGAATGAAGCAGCATACCGGCTTGCAGTTTGA 1299
DB 3071 ATCGTGAATGAAGCAGCATACCGGCTTGCAGTTTGA 3109

RESULT 2
AX042374
LOCUS AX042374 4060 bp DNA linear PAT 23-NOV-2000
DEFINITION Sequence 3 from Patent WO0064247.
ACCESSION AX042374
VERSION AX042374.1 GI:11340992
KEYWORDS
SOURCE
ORGANISM
other sequences; artificial sequences.
REFERENCE
1
AUTHORS Forsberg, C.W., Golovan, S. and Phillips, J.P.
TITLE Transgenic animals expressing salivary proteins
JOURNAL Patent: WO 0064247-A3 02-NOV-2000;
UNIVERSITY OF GUELPH (CA)
FEATURES
source
1..4060
location/Qualifiers
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="R15/AppA + Intron transgene"
ORIGIN
Query Match 100.0%; Score 1299; DB 6; Length 4060;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGAAGCATCTTAATCCATTTTATCTCTGATTCGTTAACCCGCAATCTGCA 60
DB 1811 ATGAAGCATCTTAATCCATTTTATCTCTGATTCGTTAACCCGCAATCTGCA 1870
QY 61 TTGCTCAGAGTGAAGCCGAGAGCTGAAGCTGGAAGATGTGGATGTGATGCTGATGCT 120
DB 1871 TTGCTCAGAGTGAAGCCGAGAGCTGAAGCTGGAAGATGTGGATGTGATGCTGATGCT 1930
QY 121 GTGCGTGTCCAAACCAAGGCCAAGCACTGATGACAGATGTCAACCCAGACGATGCGCA 180
DB 1931 GTGCGTGTCCAAACCAAGGCCAAGCACTGATGACAGATGTCAACCCAGACGATGCGCA 1990
QY 1931 GTGCGTGTCCAAACCAAGGCCAAGCACTGATGACAGATGTCAACCCAGACGATGCGCA 2050
DB 181 ACCTGGCGGTTAAACTGGGTTGGCTGACACCGCGGCTGGTGAATGCTATGCTATCTC 240
QY 241 GACATTAACCAAGCCAGCGCTGTGATGACCGACGATGCTGCGGCAAAAAGGCGTCCG 300
DB 2051 GACATTAACCAAGCCAGCGCTGTGATGACCGACGATGCTGCGGCAAAAAGGCGTCCG 2110
QY 301 CAGTCTGGGCAAGTGGCGGATTTATGCTGATGTGACGAGCGTAAACCGGTAAAAACAGCGCA 360
DB 2111 CAGTCTGGGCAAGTGGCGGATTTATGCTGATGTGACGAGCGTAAACCGGTAAAAACAGCGCA 2170
QY 361 GCGTTGCGCCGCGGCTGGACCTGACCTGTGCAATTAACCGTACATTAACCGGAGCATAGC 420
DB 2171 GCGTTGCGCCGCGGCTGGACCTGACCTGTGCAATTAACCGTACATTAACCGGAGCATAGC 2230
QY 421 TCCAGTCCCGATCCGTTATTTATCTCTTAAAAAATGCGGCTTGGCACTGATTAACGG 480
DB 2231 TCCAGTCCCGATCCGTTATTTATCTCTTAAAAAATGCGGCTTGGCACTGATTAACGG 2290
QY 481 AACGTGACTGACGCGATCCTCAGCAGGAGGAGGATCAATTGCTGACTTTACCGGAGCAT 540
DB 2291 AACGTGACTGACGCGATCCTCAGCAGGAGGAGGATCAATTGCTGACTTTACCGGAGCAT 2350
QY 541 CGGCAAAACGCGCTTTCGCGAATGGAACGCGGTCTTAATTTTCGCAATCAACTGTGCG 600
DB 2351 CGGCAAAACGCGCTTTCGCGAATGGAACGCGGTCTTAATTTTCGCAATCAACTGTGCG 2410
QY 601 CTTAAACGTGAGAAACAGAGCAAGAGCTGTTCAATTAACGAGCATTAACCATCGAACTC 660
DB 2411 CTTAAACGTGAGAAACAGAGCAAGAGCTGTTCAATTAACGAGCATTAACCATCGAACTC 2470
QY 661 AAGGTGAGGCGCGACCAATGCTCATTAACCGGAGGCTTAAGCGCTGATCAATGCTGAGC 720

QY 1141 CAGATGCGGATATAAAGCGCGCTGTCATTAAATACGCGCGCGGAGAGGTGAACCTGACC 1200
DB 2951 CAGATGCGGATATAAAGCGCGCTGTCATTAAATACGCGCGCGGAGAGGTGAACCTGACC 3010
QY 1201 CTGGCAGAGATGTGAAGAGCGGAATGCGCAGGCGATGTGTCGTTGGCAGGTTTATCGCA 1260
DB 3011 CTGGCAGAGATGTGAAGAGCGGAATGCGCAGGCGATGTGTCGTTGGCAGGTTTATCGCA 3070
QY 1261 ATCGTAATGAAGACGACATACCCGCTTGACGTTGTAA 1299
DB 3071 ATCGTAATGAAGACGACATACCCGCTTGACGTTGTAA 3109
RESULT 4
AX042373 6708 bp DNA linear PAT 23-NOV-2000
LOCUS AX042373 Sequence 2 from Patent WO0064247.
ACCESSION AX042373
VERSION AX042373.1 GI:11340991
KEYWORDS
SOURCE
ORGANISM
1 synthetic construct
1 synthetic construct
other sequences; artificial sequences.
REFERENCE
1 Foreberg, C.W., Golovan, S. and Phillips, J.P.
AUTHORS Transgenic animals expressing salivary proteins
TITLE Patent: WO 0064247-A 2 02-NOV-2000;
JOURNAL UNIVERSITY OF GUELPH (CA)
FEATURES
Location/Qualifiers
1..6708
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32830"
/note="R15/APPA + intron plasmid with pBLCAT3 vector"
ORIGIN
Query Match 100.0%; Score 1299; DB 6; Length 6708;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGAAAGCCATCTTAATCCCATTTTATCTCTTGATTCGTTAACCCCGCAATCTGCA 60
DB 1811 ATGAAAGCCATCTTAATCCCATTTTATCTCTTGATTCGTTAACCCCGCAATCTGCA 1870
QY 61 TTGCTCAGAGTGAAGCGGAGCTGAAAGCTGAAAGTGTGTATTTCTCAGTCTCATGCT 120
DB 1871 TTGCTCAGAGTGAAGCGGAGCTGAAAGTGTGTATTTCTCAGTCTCATGCT 1930
QY 121 GTCGCTGCTCCCAACCAAGGCGCAAGCACTGATGCGAGATGTCACCCCGCAAGCGACCA 180
DB 1931 GTCGCTGCTCCCAACCAAGGCGCAAGCACTGATGCGAGATGTCACCCCGCAAGCGACCA 1990
QY 181 ACCCTGCGCGTAAACTGGGTTTGCTGACACCGCGGCTGTGAGCTAATCCCTATCTC 240
DB 1991 ACCCTGCGCGTAAACTGGGTTTGCTGACACCGCGGCTGTGAGCTAATCCCTATCTC 2050
QY 241 GGAATTAACCAAGCGGCTGTGATGCGGACGATTCTGCGCAAAAAGGCTGCGCCG 300
DB 2051 GGAATTAACCAAGCGGCTGTGATGCGGACGATTCTGCGCAAAAAGGCTGCGCCG 2110
QY 301 CAGTCGTGTCAGAGTGCAGTATTTGCTGATGTGACGAGCGTAAACCGGTAAACAGGCGGA 360
DB 2111 CAGTCGTGTCAGAGTGCAGTATTTGCTGATGTGACGAGCGTAAACCGGTAAACAGGCGGA 2170
QY 361 GCGTTGCGCGCGCGGCTGCGACCTGATGCAATACCGTACATACCGGACAGATAGC 420
DB 2171 GCGTTGCGCGCGCGGCTGCGACCTGATGCAATACCGTACATACCGGACAGATAGC 2230
QY 421 TCCAGTCCCGATCCGTTATTTATCTCTTAATAAAGCTGGGCTTGGCACTGATTAACGCG 480
DB 2231 TCCAGTCCCGATCCGTTATTTATCTCTTAATAAAGCTGGGCTTGGCACTGATTAACGCG 2290

QY 481 AACGTACTGACGCGATCTCTCAGCAGGCGCAGAGGGTCAATTTGCTTAAACCGGGCAT 540
DB 2231 AACGTACTGACGCGATCTCTCAGCAGGCGCAGAGGGTCAATTTGCTTAAACCGGGCAT 2350
QY 541 CCGCAACCGCGGTTTCCGGAACCTGGAACGGGTGCTTAAATTTTCCGCAATCAAACTTGTGC 600
DB 2351 CCGCAACCGCGGTTTCCGGAACCTGGAACGGGTGCTTAAATTTTCCGCAATCAAACTTGTGC 2410
QY 601 CTTAAACGTGAGAAACAGAGAGAAAGCTTTCAATTAAGCGGCGCTTAAACGCTGAC 660
DB 2411 CTTAAACGTGAGAAACAGAGAGAAAGCTTTCAATTAAGCGGCGCTTAAACGCTGAC 2470
QY 661 AAGTGAGCGCGCAGCAATGTCTCATTAACCGGTGCGGTAAAGCTTCAATGCTGACG 720
DB 2471 AAGTGAGCGCGCAGCAATGTCTCATTAACCGGTGCGGTAAAGCTTCAATGCTGACG 2530
QY 721 GAGATATTTTCTCTGCAACAAGCAGAGGAATGCCGAGCCGGGGTGGGAAAGATCAC 780
DB 2531 GAGATATTTTCTCTGCAACAAGCAGAGGAATGCCGAGCCGGGGTGGGAAAGATCAC 2590
QY 781 GATTCAACCAATGGAACAACCTTGCTAATTTGCAATACGCGCAATTTTATTTGCTACAA 840
DB 2591 GATTCAACCAATGGAACAACCTTGCTAATTTGCAATACGCGCAATTTTATTTGCTACAA 2650
QY 841 CGCAGCGCAGAGGTTGCCCGCAGCGCGCCACCCCGTTATTGATTGATCAAGACAGCG 900
DB 2651 CGCAGCGCAGAGGTTGCCCGCAGCGCGCCACCCCGTTATTGATTGATCAAGACAGCG 2710
QY 901 TTGACGCCCCATCCACCGCAAAAACAGGCGTATGCTGATTAACCACTTCACTGCTG 960
DB 2711 TTGACGCCCCATCCACCGCAAAAACAGGCGTATGCTGATTAACCACTTCACTGCTG 2770
QY 961 TTTATCCGCGCAGACAGATTAATCTGCAATCTCGCGCGCGCGCACTGAGCTCAACTGG 1020
DB 2771 TTTATCCGCGCAGACAGATTAATCTGCAATCTCGCGCGCGCGCACTGAGCTCAACTGG 2830
QY 1021 ACGCTTCCCGGTCAGCGCGGATTAACAGCGCGCAGGCTGTAATCTGTTGAAACGCTGG 1080
DB 2831 ACGCTTCCCGGTCAGCGCGGATTAACAGCGCGCAGGCTGTAATCTGTTGAAACGCTGG 2890
QY 1081 CGTGGCTTAAGGATTAACAGCGGATGATTCAGGTTTCTGCTGCTTCCAGCTTTACAG 1140
DB 2891 CGTGGCTTAAGGATTAACAGCGGATGATTCAGGTTTCTGCTGCTTCCAGCTTTACAG 2950
QY 1141 CAGATGCGTATTAACCGCGCTGCTATTAATACGCCCGCGCGAGAGTGAATGTAAC 1200
DB 2951 CAGATGCGTATTAACCGCGCTGCTATTAATACGCCCGCGCGAGAGTGAATGTAAC 3010
QY 1201 CTGGCAGAGATGTGAAGAGCGAAATGCGCAGGCGATGTGTCGTTGGCAGGTTTATCGCA 1260
DB 3011 CTGGCAGAGATGTGAAGAGCGAAATGCGCAGGCGATGTGTCGTTGGCAGGTTTATCGCA 3070
QY 1261 ATCGTAATGAAGACGACATACCCGCTTGACGTTGTAA 1299
DB 3071 ATCGTAATGAAGACGACATACCCGCTTGACGTTGTAA 3109
RESULT 5
AX042378 17732 bp DNA linear PAT 23-NOV-2000
LOCUS AX042378 Sequence 7 from Patent WO0064247.
ACCESSION AX042378
VERSION AX042378.1 GI:11340996
KEYWORDS
SOURCE
ORGANISM
1 synthetic construct
1 synthetic construct
other sequences; artificial sequences.
REFERENCE
1 Foreberg, C.W., Golovan, S. and Phillips, J.P.
AUTHORS Transgenic animals expressing salivary proteins
TITLE Patent: WO 0064247-A 7 02-NOV-2000;
JOURNAL UNIVERSITY OF GUELPH (CA)
FEATURES
Location/Qualifiers

source 1. 12732
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Lama2/Alpa transgene"

Query Match 100.0%; Score 1299; DB 6; Length 17732;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 13553 TTGAGCCCCATCCACCGCAAAAGAGCGGTATGTGTGATTAACCACTTCAGTCTG 13612
Qy 961 TTTATTCGCGGACACAGATCTAATCTGCAAAATCTCGGCGGACATGAGCTCAACTGG 1020
Db 13613 TTTATTCGCGGACACAGATCTAATCTGCAAAATCTCGGCGGACATGAGCTCAACTGG 13672
Qy 1021 ACGCTTCCCGGTACGCGGATTAACAGCCCGGAGGTGTGAATGTGTGTTGAACGCTGG 1080
Db 13673 ACGCTTCCCGGTACGCGGATTAACAGCCCGGAGGTGTGAATGTGTGTTGAACGCTGG 13732
Qy 1081 CGTCGGCTAAGCGATTAACAGCAAGTGTGATTCAGTTTCGCTGCTTCCAGACTTACAG 1140
Db 13733 CGTCGGCTAAGCGATTAACAGCAAGTGTGATTCAGTTTCGCTGCTTCCAGACTTACAG 13792
Qy 1141 CAGATCCGATTAACAGCCCGCTGCTTAAATACCGCCCGGAGAGGTGAACAGACC 1200
Db 13793 CAGATCCGATTAACAGCCCGCTGCTTAAATACCGCCCGGAGAGGTGAACAGACC 13852
Qy 1201 CTGCGAGATGTGAAGAGCAAAATGCGAGGCAATGTGTTGTCAGAGTTTACGCA 1260
Db 13853 CTGCGAGATGTGAAGAGCAAAATGCGAGGCAATGTGTTGTCAGAGTTTACGCA 13912
Qy 1261 ATCGTGAATGAAGACAGCATACCCGCTTGCAGTTGTAA 1299
Db 13913 ATCGTGAATGAAGACAGCATACCCGCTTGCAGTTGTAA 13951

Db 241 GGAATTATCCAAAGCCAGCTGTGTAGCCAGCGATTTGCGGCAAAAAGGCGTCCCG 300
Qy 12893 GGAATTATCCAAAGCCAGCTGTGTAGCCAGCGATTTGCGGCAAAAAGGCGTCCCG 12952
Db 301 CAGCTGTGTAGCTGTGATTTATTTCTGATGTGACAGAGGTACCCGTTAAACAGGCGAA 360
Qy 12953 CAGCTGTGTAGCTGTGATTTATTTCTGATGTGACAGAGGTACCCGTTAAACAGGCGAA 13012
Db 361 GCGTTGCGCGCGCGGCTGAGACCGTGTGCAATACCGTACATACCCAGGCGAGATACG 420
Qy 13013 GCGTTGCGCGCGCGGCTGAGACCGTGTGCAATACCGTACATACCCAGGCGAGATACG 13072
Db 421 TCCAGTCCCGATCCGTTATTTAATCTCTTAAACCTGCGCTTCCCACTGATACGCG 480
Qy 13073 TCCAGTCCCGATCCGTTATTTAATCTCTTAAACCTGCGCTTCCCACTGATACGCG 13132
Db 481 AACGTGACTAGCGGATCTCAGCAGGCGAGAGGGTCAATTGCTGACTTTACCGGCGAT 540
Qy 13133 AACGTGACTAGCGGATCTCAGCAGGCGAGAGGGTCAATTGCTGACTTTACCGGCGAT 13192
Db 541 CGGCAAAAGCGCGTTTCCGCAACCTGGAACGGGTGCTTAATTTCCGCAATCAAACTGTGC 600
Qy 13193 CGGCAAAAGCGCGTTTCCGCAACCTGGAACGGGTGCTTAATTTCCGCAATCAAACTGTGC 13252
Db 601 CTTAAACGTGAGAAACAGAGCGAAGCTGTTCATTAAACGAGCATTAACATCGGAATC 660
Qy 13253 CTTAAACGTGAGAAACAGAGCGAAGCTGTTCATTAAACGAGCATTAACATCGGAATC 13312
Db 661 AAGGTGAGCGCGCAATGTCTCTTAAACCGGTGCGGTAAAGCTTCGATCAATGCTGACG 720
Qy 13313 AAGGTGAGCGCGCAATGTCTCTTAAACCGGTGCGGTAAAGCTTCGATCAATGCTGACG 13372
Db 721 GAGATATTTCTCTCTGCAACAGCAACAGGAAATGCGGAGCGCGGTGAGGGAAGATACG 780
Qy 13373 GAGATATTTCTCTCTGCAACAGCAACAGGAAATGCGGAGCGCGGTGAGGGAAGATACG 13432
Db 781 GATTCAACAGAGTGAACACTTGTCTAAGTTTGCATTAACGCGCAATTTATTTCTCTCAA 840
Qy 13433 GATTCAACAGAGTGAACACTTGTCTAAGTTTGCATTAACGCGCAATTTATTTCTCTCAA 13492
Db 841 CGCAGCGCAAGGTTTCCCGCAGCGCGCGCAATTTATTAAGTTTGAATTAAGACAGCG 900
Qy 13493 CGCAGCGCAAGGTTTCCCGCAGCGCGCGCAATTTATTAAGTTTGAATTAAGACAGCG 13552
Db 901 TTGAGCGCCCATTCACCGCAAAACAGGCGTATGTGTGATTAACCACTTCAGTGTG 960

Db 13553 TTGAGCCCCATCCACCGCAAAAGAGCGGTATGTGTGATTAACCACTTCAGTCTG 13612
Qy 961 TTTATTCGCGGACACAGATCTAATCTGCAAAATCTCGGCGGACATGAGCTCAACTGG 1020
Db 13613 TTTATTCGCGGACACAGATCTAATCTGCAAAATCTCGGCGGACATGAGCTCAACTGG 13672
Qy 1021 ACGCTTCCCGGTACGCGGATTAACAGCCCGGAGGTGTGAATGTGTGTTGAACGCTGG 1080
Db 13673 ACGCTTCCCGGTACGCGGATTAACAGCCCGGAGGTGTGAATGTGTGTTGAACGCTGG 13732
Qy 1081 CGTCGGCTAAGCGATTAACAGCAAGTGTGATTCAGTTTCGCTGCTTCCAGACTTACAG 1140
Db 13733 CGTCGGCTAAGCGATTAACAGCAAGTGTGATTCAGTTTCGCTGCTTCCAGACTTACAG 13792
Qy 1141 CAGATCCGATTAACAGCCCGCTGCTTAAATACCGCCCGGAGAGGTGAACAGACC 1200
Db 13793 CAGATCCGATTAACAGCCCGCTGCTTAAATACCGCCCGGAGAGGTGAACAGACC 13852
Qy 1201 CTGCGAGATGTGAAGAGCAAAATGCGAGGCAATGTGTTGTCAGAGTTTACGCA 1260
Db 13853 CTGCGAGATGTGAAGAGCAAAATGCGAGGCAATGTGTTGTCAGAGTTTACGCA 13912
Qy 1261 ATCGTGAATGAAGACAGCATACCCGCTTGCAGTTGTAA 1299
Db 13913 ATCGTGAATGAAGACAGCATACCCGCTTGCAGTTGTAA 13951

RESULT 6
AX042372 20623 bp DNA linear PAT 23-NOV-2000
LOCUS
DEFINITION Sequence 1 from Patent WO0064247.
ACCESSION AX042372
VERSION AX042372.1 GI:11340990
KEYWORDS
SOURCE
ORGANISM
synthetic construct
synthetic construct
other sequences; artificial sequences.

REFERENCE
1 Forsberg, C.W., Golovan, S. and Philippe, J.P.
Transgenic animals expressing salivary proteins
Patent: WO 0064247-A 1 02-NOV-2000;
JOURNAL
UNIVERSITY OF GUELPH (CA)
location/Qualifiers

FEATURES
source
1. 20623
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Lama2/Alpa plasmid"

ORIGIN

Query Match 100.0%; Score 1299; DB 6; Length 20623;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 13553 TTGAGCCCCATCCACCGCAAAAGAGCGGTATGTGTGATTAACCACTTCAGTCTG 13612
Qy 961 TTTATTCGCGGACACAGATCTAATCTGCAAAATCTCGGCGGACATGAGCTCAACTGG 1020
Db 13613 TTTATTCGCGGACACAGATCTAATCTGCAAAATCTCGGCGGACATGAGCTCAACTGG 13672
Qy 1021 ACGCTTCCCGGTACGCGGATTAACAGCCCGGAGGTGTGAATGTGTGTTGAACGCTGG 1080
Db 13673 ACGCTTCCCGGTACGCGGATTAACAGCCCGGAGGTGTGAATGTGTGTTGAACGCTGG 13732
Qy 1081 CGTCGGCTAAGCGATTAACAGCAAGTGTGATTCAGTTTCGCTGCTTCCAGACTTACAG 1140
Db 13733 CGTCGGCTAAGCGATTAACAGCAAGTGTGATTCAGTTTCGCTGCTTCCAGACTTACAG 13792
Qy 1141 CAGATCCGATTAACAGCCCGCTGCTTAAATACCGCCCGGAGAGGTGAACAGACC 1200
Db 13793 CAGATCCGATTAACAGCCCGCTGCTTAAATACCGCCCGGAGAGGTGAACAGACC 13852
Qy 1201 CTGCGAGATGTGAAGAGCAAAATGCGAGGCAATGTGTTGTCAGAGTTTACGCA 1260
Db 13853 CTGCGAGATGTGAAGAGCAAAATGCGAGGCAATGTGTTGTCAGAGTTTACGCA 13912
Qy 1261 ATCGTGAATGAAGACAGCATACCCGCTTGCAGTTGTAA 1299
Db 13913 ATCGTGAATGAAGACAGCATACCCGCTTGCAGTTGTAA 13951

LOCUS	COORDINATES	SEQUENCE	LENGTH
Db	301	CAGCTCTGCTCAGAGTCCGCAATTAATGGTGAATGTCGACAGACCGTACCCGTAACAAACAGGCGAA	360
	12953	CAGCTCTGCTCAGAGTCCGCAATTAATGGTGAATGTCGACAGACCGTACCCGTAACAAACAGGCGAA	13012
	361	GCCTTCGCGCCCGGGGCTGGCACCCTGCACTGTGCAATTAACCGTACATACCAGGCAATACG	420
	13013	GCCTTCGCGCCCGGGGCTGGCACCCTGCACTGTGCAATTAACCGTACATACCAGGCAATACG	13072
	421	TCCAGTCCCGAATCCGTTATTAAATCCTCTTAAAACTGGGGTTTGGCAACTGGATTAACGCG	480
	13073	TCCAGTCCCGAATCCGTTATTAAATCCTCTTAAAACTGGGGTTTGGCAACTGGATTAACGCG	13132
	481	AACGTGACTACGCGATCCTCAGCAGGGCAGAGGGCTCAATTGCTGACCTTTACCGGGCAT	540
	13133	AACGTGACTACGCGATCCTCAGCAGGGCAGAGGGCTCAATTGCTGACCTTTACCGGGCAT	13192
	541	CGGCAAAACGCGGTTTCGCGAATGGAACGGGCTTAAATTTTCGCAATCAAACTTGTGC	600
	13193	CGGCAAAACGCGGTTTCGCGAATGGAACGGGCTTAAATTTTCGCAATCAAACTTGTGC	13252
601	CTTAAACGTGGAACACGACGAAAGCTGTTCAATTAACGAGGCAATTAACATCCGAACTC	660	
13253	CTTAAACGTGGAACACGACGAAAGCTGTTCAATTAACGAGGCAATTAACATCCGAACTC	13312	
661	AAGGTGAGCGCGGCAATGTCCTAATTAACGCGGCGGTAAAGCTCGCATCAATGCTGACG	720	
13313	AAGGTGAGCGCGGCAATGTCCTAATTAACGCGGCGGTAAAGCTCGCATCAATGCTGACG	13372	
721	GAGATATTTCCTCTGCAACAAAGCAACAGGAATCCGAGCCGCGGGTGGGGAAGATCACC	780	
13373	GAGATATTTCCTCTGCAACAAAGCAACAGGAATCCGAGCCGCGGGTGGGGAAGATCACC	13432	
781	GATTCAACACGAGTGAACAACCTTCTAAGTTTGATTAACGCGAATTTATTATTGCTACAA	840	
13433	GATTCAACACGAGTGAACAACCTTCTAAGTTTGATTAACGCGAATTTATTATTGCTACAA	13492	
841	CGCAGCGCAGAGGTGGCCCGCAGCGCGCGCACCCCGTTATTGAATTTGATCAAGACAGCG	900	
13493	CGCAGCGCAGAGGTGGCCCGCAGCGCGCGCACCCCGTTATTGAATTTGATCAAGACAGCG	13552	
901	TTGACGCGCCCATCCACCGCAAAAAACAGCGTATGGTGTGACATTAACCACTTAGTGCTG	960	
13553	TTGACGCGCCCATCCACCGCAAAAAACAGCGTATGGTGTGACATTAACCACTTAGTGCTG	13612	
961	TTTATACGCCGACACGATTAATCTGGCAAACTCGCGCGCGCACGACGAGCTCAACTGG	1020	
13613	TTTATACGCCGACACGATTAATCTGGCAAACTCGCGCGCGCACGACGAGCTCAACTGG	13672	
1021	AACGCTCCCGGTCACCGGATTAACACGCGCCAGAGTGGTGAACAGGTTTGAACGCTGG	1080	
13673	AACGCTCCCGGTCACCGGATTAACACGCGCCAGAGTGGTGAACAGGTTTGAACGCTGG	13732	
1081	CGTCCGCTAAAGCATPAACGACGATGATTAAGGTTTGCCTGGCTTCCAGACTTTAACG	1140	
13733	CGTCCGCTAAAGCATPAACGACGATGATTAAGGTTTGCCTGGCTTCCAGACTTTAACG	13792	
1141	CAGATCCGCGATTAACGCGGTGCTCATTAATACGCGCCCGGAGAGTGAACATGACC	1200	
13793	CAGATCCGCGATTAACGCGGTGCTCATTAATACGCGCCCGGAGAGTGAACATGACC	13852	
1201	CTGGCAGAGTGTGAAGAGCGAANTCGACAGGCGATGTTGTTGGCAGGTTTTCGCA	1260	
13853	CTGGCAGAGTGTGAAGAGCGAANTCGACAGGCGATGTTGTTGTTGGCAGGTTTTCGCA	13912	
1261	ATCGATGATGAAGACGACGATACCCGCTTGCACTTTGTAA	1299	
13913	ATCGATGATGAAGACGACGATACCCGCTTGCACTTTGTAA	13951	

DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE
...

REFERENCE
AUTHORS

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

TITLE
JOURNAL
REFERENCE

AUTHORS
TITLE
JOURNAL

COMMENT

Escherichia coli K12 genomic DNA. (22.3 - 22.7 min).
D90733.1 AB001340
D90733.1 GI:4062539
Complete and shotgun sequencing; hyaB; hyaC; hyaD; hyaE; hyaF;
appC; cyxB; cbbA; appB; cyxB; cbbB; appA; amsI; insA; insB; cspF;
cspB; sfa; yccL; yccM.
Escherichia coli K12
Escherichia coli K12
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacterales;
Enterobacteriaceae; Escherichia.

1
Oshima, T., Alba, H., Baba, T., Fujita, K., Hayaishi, K., Honjo, A.,
Ikemoto, K., Inada, T., Itoh, T., Kajihara, M., Kanai, K., Kashiwano, K.,
Kimura, S., Kitagawa, M., Makino, K., Masuda, S., Miki, T.,
Mizobuchi, K., Mori, H., Motomura, K., Nakamura, Y., Nishimoto, H.,
Nishio, Y., Saito, N., Sampei, G., Seki, Y., Tagami, H., Takemoto, K.,
Wada, C., Yamamoto, Y., Yano, M. and Horuchi, T.
A 718-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map
DNA Res. 3 (3), 137-155 (1996)
97061202
8905232

2
Alba, H., Baba, T., Fujita, K., Hayaishi, K., Honjo, A., Horuchi, T.,
Ikemoto, K., Inada, T., Isono, S., Isono, S., Itoh, T., Kanai, K.,
Kasai, H., Kashiwano, K., Kim, S., Kimura, S., Kitagawa, M.,
Kitakawa, M., Makino, K., Masuda, S., Miki, T., Mizobuchi, K., Mori, H.,
Motomura, K., Nakamura, Y., Nishimoto, H., Nishio, Y., Oshima, T.,
Saito, N., Sampei, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C.,
Yamamoto, Y. and Yano, M.
The systematic sequencing of the Escherichia coli genome in Japan
Unpublished
3 (bases 1 to 18882)

3
Mori, H.
Direct Submission
Submitted (29-JUL-1996) Hirofada Mori, NARA Institute of Science
and Technology, Res. & Edu. Center for Genetic Info., 8916-5
Takayama, Ikoma, Nara 630-01, Japan
(E-mail:hmori@gtc.aisf-nara.ac.jp), Tel:81-7437-2-5660,
Fax:81-7437-2-5669)
On Dec 24, 1998 this sequence version replaced gi:1651473.
Project: Enterobacter Information:
The Japan E.coli genome DNA sequencing project
Group:

FEATURES

```
URL:
The Japan E. coli genome database
http://bsw3.aist-nara.ac.jp.
Location/Qualifiers
1..16882
/organism="Escherichia coli K12"
/mol_type="genomic DNA"
/strain="K12"
/db_xref="taxon:83333"
```

```

/organism="Escherichia coli K12"
/mol_type="genomic DNA"
/strain="K12"
/db_xref="taxon:83333"

```


/map="22.3-22.7 min"
/clone="Kohara clone #225"
/note="Nucleotide position 1034630-1053511 from the
initiation site of Thra (0 min). This clone is from
Kohara lambda miniset library."
1. .966
/gene="hyab"
CDS
/gene="hyab"
/note="ORF ID:0224#6
similar to SwissProt Accession Number P19927"
/codon_start=1
/transl_table=11
/product="Hydrogenase-1 large chain (EC 1.18.99.1) (nife
hydrogenase) (membrane-bound hydrogenase 1 large subunit)
(hyd1)."
/protein_id="BAA35738.1"
/db_xref="GI:4062540"
/translation="PDALAIAGQFNKPKSEIGTGLSDKCVLSGAPPDIANRGEKSL
MGGAIVNGFNVLVDVDPQVDFVDHARYRPNDVGRHPFDGTTDPVYRNGD
VKGSDNIQOLNEOERYSWIKAPRMGNAMVEGPLARTLIAHKGDAIVESVDKMS
ALNLPLSGIQTSLGRILCRHAEQAQAGLQYFPDKLMTNLKNGNLATISTEKMBAT
WPTCEGVEFTBAPRGALGHMAALRDGKIDLYOCVPTTNASPRDPKQIGAYEAL
MTKMAIPEQPLEILRLTHSFDPCLACSTHVLGDSDSELISQVR"
985. .1692
/gene="hyac"
CDS
985. .1692
/gene="hyac"
/note="ORF ID:0225#1
similar to PIR Accession Number JV0074"
/codon_start=1
/transl_table=11
/product="Hyac protein"
/protein_id="BAA35739.1"
/db_xref="GI:1651475"
/translation="MOCKSDNVSHVREAPVRIMHLLTVLCMAVLMVTCYFICLP
SVSGEATLTFMGITLRLHPSAGNVFTVILMKIYAFVGNRRISRELFITPVKSKM
QGVWEIRWYLFLAKRPSADIGHNPIAQAMFGFGLVSMITGFALYSHEQYALF
APRYVVEFPYTCGNSMDHSMHRLGMVLIGAFVIGHVYMALREDIMSDVTSTW
NGYSHKFKGISNKERS"
1689. .2276
/gene="hyad"
CDS
1689. .2276
/gene="hyad"
/note="ORF ID:0225#2
similar to SwissProt Accession Number P19930"
/codon_start=1
/transl_table=11
/product="Hydrogenase-1 operon protein hyad."
/protein_id="BAA35740.1"
/db_xref="GI:1651476"
/translation="MSEORVVMGLGNILNADBEQGVAVARELYAHYHMBEVEIYDG
GTGGINLGVESASHLILDAIDYGLBPGTLRTYAGERIPAYLSAKMGLHONSFE
VLAADIRGHLPAHIALVGLQPMALDYGSGLSLBAEQLPAEQALLAQLANGIYF
OPAESRCLNVDCLSMENYEGVRLRYRMTQEBG"
2273. .2671
/gene="hyab"
CDS
2273. .2671
/gene="hyab"
/note="ORF ID:0225#3
similar to SwissProt Accession Number P19931"
/codon_start=1
/transl_table=11
/product="Hydrogenase-1 operon protein hyab."
/protein_id="BAA35741.1"
/db_xref="GI:1651477"
/translation="MSNDTFPDALMQMLARGTTPVESRLDMLTLPAPDGVLLSD
PKRTPEVDNPMVIGELRERPDYTQVVALADLEQSAIGDRFGVFRFPATLVFTGGA
YRGVINGIHVPAEILINLMRGLVEPQERAS"
2668. .3525
/gene="hyaf"
CDS
2668. .3525
/gene="hyaf"

/note="ORF ID:0225#4
similar to SwissProt Accession Number P19932"
/codon_start=1
/transl_table=11
/product="Hydrogenase-1 operon protein hyaf."
/protein_id="BAA35742.1"
/db_xref="GI:1651478"
/translation="MSETFHLLSGPTQPNDSFSMNPPLITCOVNDSPMALEQCA
HSPQVYALINEHQHLSERQPRIGVLAVDILNADDDHFIINTLLGEGEVRIQOA
DSESEIODEAIFCGLMVRKRGESKLEKLEAGCAPLALMQAATONLLPTSLRPP
IDGLNKGLEPLAEHLAHVNPDAQPHSINTQLPISADRLFSRLCGGENTQIRITG
YGESYINATGLRHVHLRCTDTLKGPLESYBICPIPEVVALAPEDVDSADRLSEVC
OMLAEPAP"
3659. .5203
/gene="appc"
/note="synonyms: cbda, cyxa"
3659. .5203
/gene="appc"
/note="ORF ID:0225#5
similar to SwissProt Accession Number P26459"
/codon_start=1
/transl_table=11
/product="Cytochrome bd-ii oxidase subunit i (EC
1.10.3.-)."
/protein_id="BAA35743.1"
/db_xref="GI:1651479"
/translation="MNDVIDLSRWQPALTKHFLFVPLTGLIFLILAMERTVYVTC
KITIYRDMTRFWGKLFGINFALGATGLTNEPFGINTWSYFSNVGIFGAPPLAMEALM
AFLESTFVGLFEFGQWRLNKYQHLVTVLVAFGSLSLMTLINANMQPYGAHFD
IDTLRMBMTSFSELVNPVSOYVFTVMAQVYTCAMFIMAIISAWLLGRERNVLR
SPAIGSPGTGLIITGLIOLGDSAYEAVQVOPVYKLAAMEGMEQTEBPAPFHVAMP
QDQERNAFLAKTIPALLGITATISLSDKPVGICIKIMAEYTPRLQGRMVLIMQEI
NEBPVLAQFRGLBDDGGLMSLRYPAMNHTTAQYQAARKGAIPOVAPVNSPRI
MGGCSLLILVMIILAVQTLRKLIDQHRVVLKVALSLPLVTAIAGPMFIEFGOP
WAIDILPTYSASHALTQQLAFSLIMIVGLYTLFLAEVYLMQKXARLGRPSAMQSO
PQQOQ"
5215. .6351
/gene="appb"
/note="synonyms: cbdb, cyxb"
5215. .6351
/gene="appb"
/note="ORF ID:0225#6
similar to SwissProt Accession Number P26458"
/codon_start=1
/transl_table=11
/product="Cytochrome bd-ii oxidase subunit ii (EC
1.10.3.-)."
/protein_id="BAA35744.1"
/db_xref="GI:1651480"
/translation="MPDYETLRFTIMWLLITGVILVPMISGPFMGICULLPIVARND
ERRIVINSVGAHNEGQVWILILGALFLAMPRVYAAASGFTVAMILVICSIFPRPL
AFPRKIKIDARRRKMDAGLVIGSLVPVVFIRAGNLLGVFPAPFQPLRVEYLS
FMQLLHPFPLICGLSLGWIYLOGGVWLQIKTVGVIHLRSQALTKKAAALLVMLCFLLA
Query Match 99.6%; Score 1294.2; DB 1; Length 18882;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1296; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATGAAGCATCTTAATCCATTTTATCTCTTCTGATTCCTGTAACCCCGCATATCGCA 60
DB 6536 ATGAAGCATCTTAATCCATTTTATCTCTTCTGATTCCTGTAACCCCGCATATCGCA 6595
QY 61 TTGCTCAGAGTGAAGCCGAGCTGAAGCTGGAAGTGTGTGATTTGTCAATGCTCATGCT 120
DB 6596 TTGCTCAGAGTGAAGCCGAGCTGAAGCTGGAAGTGTGTGATTTGTCAATGCTCATGCT 6655
QY 121 GTGCTGCTCAACCAAGGCGACGCAACGATGAGGATGCAACCCGACGACGATGCGCA 180
DB 6656 GTGCTGCTCAACCAAGGCGACGCAACGATGAGGATGCAACCCGACGACGATGCGCA 6715
QY 181 ACCTGCGCGTAAACTGGTGTGCTGACACCGCGCGGTGTGATGCTTAATCGCTATCTC 240
DB 6716 ACCTGCGCGTAAACTGGTGTGCTGACACCGCGCGGTGTGATGCTTAATCGCTATCTC 6775

QY 241 GCACATTACCAAGCCAGCGCTGTGGTAGCCGACGGAATTCCTGGCCGAAAAAGGCGTCCCG 300
 DB 6776 GGAACATTACCAAGCCAGCGCTGTGGTAGCCGACGGAATTCCTGGCCGAAAAAGGCGTCCCG 6835
 QY 301 CAGTCGTGACGAGTGGCGGATTAATTCGATGTCGACGAGCGTACCCTGTAACAGCGGCA 360
 DB 6836 CAGTCGTGACGAGTGGCGGATTAATTCGATGTCGACGAGCGTACCCTGTAACAGCGGCA 6895
 QY 361 GCCTTCGCGCGCGGCGTGGCACTGTCGATTAACCGTACATACCAGGCGAGATACG 420
 DB 6896 GCCTTCGCGCGCGGCGTGGCACTGTCGATTAACCGTACATACCAGGCGAGATACG 6955
 QY 421 TCCAGTCCCGATCCGTTATTTAACTCTTAATAAACTGGGTTTGCCCACTGGATAACGG 480
 DB 6956 TCCAGTCCCGATCCGTTATTTAACTCTTAATAAACTGGGTTTGCCCACTGGATAACGG 7015
 QY 481 AACGTGACTGAGCGGATCCCTGACGAGGAGAGGAGGCTCAATTGCTGACTTTACCGGGCAT 540
 DB 7016 AACGTGACTGAGCGGATCCCTGACGAGGAGAGGAGGCTCAATTGCTGACTTTACCGGGCAT 7075
 QY 541 CGGCAAAACGCGCTTTGCGCACTGGAACGCGTGTCTTAATTTTCCGATCAAACTTGTCG 600
 DB 7076 CGGCAAAACGCGCTTTGCGCACTGGAACGCGTGTCTTAATTTTCCGATCAAACTTGTCG 7135
 QY 601 CTTAAACGTGAGAAACAGGACGAACTGTTCATTAACGAGGCACTTAACATCGGAACCTC 660
 DB 7136 CTTAAACGTGAGAAACAGGACGAACTGTTCATTAACGAGGCACTTAACATCGGAACCTC 7195
 QY 661 AAGGTGAGCGCGCAATGTCCTATTAAACGAGTGCAGTAAAGCTCCGATCAATGCTGACG 720
 DB 7196 AAGGTGAGCGCGCAATGTCCTATTAAACGAGTGCAGTAAAGCTCCGATCAATGCTGACG 7255
 QY 721 GAGATATTTCTCTGCAACAGCAACAGGATTCGGAACCGGCGTGGGGAGAGATCACG 780
 DB 7256 GAGATATTTCTCTGCAACAGCAACAGGATTCGGAACCGGCGTGGGGAGAGATCACG 7315
 QY 781 GATTTCACAGAGGGAACAACCTTGTAAGTTGATTAACGCAATTTTATTTGCTACAA 840
 DB 7316 GATTTCACAGAGGGAACAACCTTGTAAGTTGATTAACGCAATTTTATTTGCTACAA 7375
 QY 841 CGCAGCGCAGAGGTTGCCGACGCGCGCCACCCGTTATTAGATTGATCAAGACGCG 900
 DB 7376 CGCAGCGCAGAGGTTGCCGACGCGCGCCACCCGTTATTAGATTGATCAAGACGCG 7435
 QY 901 TTGAACGCCCATCCACCGCAAAAACAGCGGTAATGTTGATCAATCCACTTCACTGCTG 960
 DB 7436 TTGAACGCCCATCCACCGCAAAAACAGCGGTAATGTTGATCAATCCACTTCACTGCTG 7495
 QY 961 TTTATCGCGGGAACGATCTAATCTGCAAAATCTCGGCGGCGCACTGAGCTCAACTGG 1020
 DB 7496 TTTATCGCGGGAACGATCTAATCTGCAAAATCTCGGCGGCGCACTGAGCTCAACTGG 7555
 QY 1021 ACGCTTCCCGGTACGCGGATTAACGCGCGCAAGGTGTGAATCTGTTTGAACGCTGG 1080
 DB 7556 ACGCTTCCCGGTACGCGGATTAACGCGCGCAAGGTGTGAATCTGTTTGAACGCTGG 7615
 QY 1081 CGTGGCTAAGCCATTAACGCACTGATTCAGGTTTCGTGCTTCCGACTTAAACG 1140
 DB 7616 CGTGGCTAAGCCATTAACGCACTGATTCAGGTTTCGTGCTTCCGACTTAAACG 7675
 QY 1141 CAGATGCGGTATAAAAGCGGCTGTCAATTAATACGCGCGCGGAGAGGTGAACCTGACC 1200
 DB 7676 CAGATGCGGTATAAAAGCGGCTGTCAATTAATACGCGCGCGGAGAGGTGAACCTGACC 7735
 QY 1201 CTGGCAGATGTGAAGAGCGAAATGCGCAAGGCGATGTTGCTGTTGCGAGGTTTACGCA 1260
 DB 7736 CTGGCAGATGTGAAGAGCGAAATGCGCAAGGCGATGTTGCTGTTGCGAGGTTTACGCA 7795
 QY 1261 ATGCTGAATGAAGACGATACCGGCTTGCACTTTTGTAA 1299
 DB 7796 ATGCTGAATGAAGACGATACCGGCTTGCACTTTTGTAA 7834

RESULT 8
 U0096.10
 WPCOMMENT
 Sequence split into 47 fragments LOCUS U0096 Accession U0096

Fragment Name	Begin	End
U0096_00	1	110000
U0096_01	100001	210000
U0096_02	200001	310000
U0096_03	300001	410000
U0096_04	400001	510000
U0096_05	500001	610000
U0096_06	600001	710000
U0096_07	700001	810000
U0096_08	800001	910000
U0096_09	900001	1010000
U0096_10	1000001	1110000
U0096_11	1100001	1210000
U0096_12	1200001	1310000
U0096_13	1300001	1410000
U0096_14	1400001	1510000
U0096_15	1500001	1610000
U0096_16	1600001	1710000
U0096_17	1700001	1810000
U0096_18	1800001	1910000
U0096_19	1900001	2010000
U0096_20	2000001	2110000
U0096_21	2100001	2210000
U0096_22	2200001	2310000
U0096_23	2300001	2410000
U0096_24	2400001	2510000
U0096_25	2500001	2610000
U0096_26	2600001	2710000
U0096_27	2700001	2810000
U0096_28	2800001	2910000
U0096_29	2900001	3010000
U0096_30	3000001	3110000
U0096_31	3100001	3210000
U0096_32	3200001	3310000
U0096_33	3300001	3410000
U0096_34	3400001	3510000
U0096_35	3500001	3610000
U0096_36	3600001	3710000
U0096_37	3700001	3810000
U0096_38	3800001	3910000
U0096_39	3900001	4010000
U0096_40	4000001	4110000
U0096_41	4100001	4210000
U0096_42	4200001	4310000
U0096_43	4300001	4410000
U0096_44	4400001	4510000
U0096_45	4500001	4610000
U0096_46	4600001	4639675

 Continuation (11 of 47) of U0096 from base 1000001 (U0096 Escherichia coli K-12 MG1655
 Query Match 99.6%; Score 1294.2; DB 1; Length 110000;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1296; Mismatches 0; Indels 3; Gaps 0;

QY 1 ATGAAGCATCTTAATCCATTTTATCTCTGATTCGTTAACCCTGCAATCTGCA 60
 DB 39840 ATGAAGCATCTTAATCCATTTTATCTCTGATTCGTTAACCCTGCAATCTGCA 39899
 QY 61 TTGCTCAGAGTAGCGCGAGCTGGAAGTGAAGTGTGATTTGTCAGTGCATGCT 120
 DB 39900 TTGCTCAGAGTAGCGCGAGCTGGAAGTGAAGTGTGATTTGTCAGTGCATGCT 39959
 QY 121 GTGCGTCTCCAAACCAAGCCACGCACTGATGCAAGATGTACCCCAAGCGATGGCCA 180
 DB 39960 GTGCGTCTCCAAACCAAGCCACGCACTGATGCAAGATGTACCCCAAGCGATGGCCA 40019
 QY 181 ACTGCGCGGTAAACTGGGTTGGCTGACACCGCGGCGTGAAGCTAATGCGCTATCTC 240

Db 40020 ACCGCGCGGTAAGAACTGGGTTGGCTGACACCGCGCGGTGCTGAGCTAATCCGCTATCTC 40079
 QY 241 GGCATTTACCAACGCGACGCTGTGTAGCCGACGATGCTGCGAAAAAGGCTGCCG 300
 Db 40080 GGCATTTACCAACGCGACGCTGTGTAGCCGACGATGCTGCGAAAAAGGCTGCCG 40139
 QY 301 CAGCTGTGTCAGGTCGCGATTTATGCTGATGTGACGAGGCTACCCGTAAACAGCGAA 360
 Db 40140 CAGCTGTGTCAGGTCGCGATTTATGCTGATGTGACGAGGCTACCCGTAAACAGCGAA 40199
 QY 361 GCCTTGGCGCGCGGGGCTGGACCTGACTGCAATACCGTACATACCAGGAGATACG 420
 Db 40200 GCCTTGGCGCGCGGGGCTGGACCTGACTGCAATACCGTACATACCAGGAGATACG 40259
 QY 421 TCCAGTCCCGATCCGTTATTTAATCCTTTAAAACTGGCGCTTTGCCAATGATTAACGCG 480
 Db 40260 TCCAGTCCCGATCCGTTATTTAATCCTTTAAAACTGGCGCTTTGCCAATGATTAACGCG 40319
 QY 481 AAGTGAATGACCGCATCTTCACACGAGGCGAGGGGTCAATTCGATCTTACCGGGCAT 540
 Db 40320 AAGTGAATGACCGCATCTTCACACGAGGCGAGGGGTCAATTCGATCTTACCGGGCAT 40379
 QY 541 CGGCAACCGCGCTTTGCGCAACTGGAACGGGTGCTTAATTTTCCGCAATCAAACTTGTGC 600
 Db 40380 CGGCAACCGCGCTTTGCGCAACTGGAACGGGTGCTTAATTTTCCGCAATCAAACTTGTGC 40439
 QY 601 CTTAAACGTGAGAAACAGACGAAAGCTGTTCAATTAACGAGGATTAACATCGAACTC 660
 Db 40440 CTTAAACGTGAGAAACAGACGAAAGCTGTTCAATTAACGAGGATTAACATCGAACTC 40499
 QY 661 AAGTGAAGCCCGCAATGTCTCATTTAACGGGTGCGGTAAAGCTCCGATCAATGCTGACG 720
 Db 40500 AAGTGAAGCCCGCAATGTCTCATTTAACGGGTGCGGTAAAGCTCCGATCAATGCTGACG 40559
 QY 721 GAGATATTTCTCTGCAACAGACAGGGAATGCCGAGCCGGGTGGGAGGATACCC 780
 Db 40560 GAGATATTTCTCTGCAACAGACAGGGAATGCCGAGCCGGGTGGGAGGATACCC 40619
 QY 781 GATTCAACCAAGTGAACACCTTGCTAAGTTTGCAATGCGCAATTTTAATTTGCTACAA 840
 Db 40620 GATTCAACCAAGTGAACACCTTGCTAAGTTTGCAATGCGCAATTTTAATTTGCTACAA 40679
 QY 841 CGCACGCGCAAGGTGCTCCGCGACGCGCGCCACCCGCTTATTAATTTGATCAACAGCGG 900
 Db 40680 CGCACGCGCAAGGTGCTCCGCGACGCGCGCCACCCGCTTATTAATTTGATCAACAGCGG 40739
 QY 901 TTGACGCGCCATCCACGCGCAAAAACAGGCGATGCTGATCAATACCACTTCAGTCTG 960
 Db 40740 TTGACGCGCCATCCACGCGCAAAAACAGGCGATGCTGATCAATACCACTTCAGTCTG 40799
 QY 961 TTATTCGCGGACAGATACTAATCTGCAAAATCTCGCGCGCGCATGGAAGCTCACTG 1020
 Db 40800 TTATTCGCGGACAGATACTAATCTGCAAAATCTCGCGCGCGCATGGAAGCTCACTG 40859
 QY 1021 ACGCTTCCCGGTACGCGGATTAACGCGCGCAAGTGTGAACGCTGTTGAACGCTGG 1080
 Db 40860 ACGCTTCCCGGTACGCGGATTAACGCGCGCAAGTGTGAACGCTGTTGAACGCTGG 40919
 QY 1081 CGTGGGCTAAGCGATTAACAGCAGTGAATCAGGTTTCGCTGCTTCACACTTTACAG 1140
 Db 40920 CGTGGGCTAAGCGATTAACAGCAGTGAATCAGGTTTCGCTGCTTCACACTTTACAG 40979
 QY 1141 CAGATGCGGTATTAACGCGCGCTGTCAATTAATACGCGCGCGGAGAGGTGAACCTGACC 1200
 Db 40980 CAGATGCGGTATTAACGCGCGCTGTCAATTAATACGCGCGCGGAGAGGTGAACCTGACC 41039
 QY 1201 CTGGCAGGATGTGAAGAGCGAAATGCGGAGGCGATGTGTTGGCGAGTTTACGAA 1260
 Db 41040 CTGGCAGGATGTGAAGAGCGAAATGCGGAGGCGATGTGTTGGCGAGTTTACGAA 41099
 QY 1261 ATCGGAATGAAGACGCGATACCGGCTTGAAGTTTGTAA 1299
 Db 41100 ATCGGAATGAAGACGCGATACCGGCTTGAAGTTTGTAA 41138

RESULT 9
 ECOAPPAA 1901 bp DNA linear BCT 04-APR-2002
 LOCUS Escherichia coli periplasmic phosphoanhydride phosphohydrolase
 DEFINITION (appa) gene, complete cds.
 ACCESSION M58708.1 GI:145283
 VERSION M58708.1
 KEYWORDS Escherichia coli
 SOURCE Escherichia coli
 ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
 REFERENCE 1 (bases 1 to 1901)
 AUTHORS Dases, J., Marck, C., and Boquet, P.L.
 TITLE The complete nucleotide sequence of the Escherichia coli gene appa reveals significant homology between pH 2.5 acid phosphatase and glucose-1-phosphatase
 JOURNAL J. Bacteriol. 172 (9), 5497-5500 (1990)
 MEDLINE 90368616
 PUBMED 2168385
 REFERENCE 2 (sites)
 AUTHORS Oestelin, K., Harms, E.H., Stevis, P.E., Kuciel, R., Zhou, M.M. and Van Etten, R.L.
 TITLE Overexpression, site-directed mutagenesis, and mechanism of Escherichia coli acid phosphatase
 JOURNAL J. Biol. Chem. 267 (32), 22830-22836 (1992)
 MEDLINE 93054596
 PUBMED 1429631
 REFERENCE 3 (sites)
 AUTHORS Oestelin, K. and Van Etten, R.L.
 TITLE Asp304 of Escherichia coli acid phosphatase is involved in leaving group protonation
 JOURNAL J. Biol. Chem. 268 (28), 20778-20784 (1993)
 MEDLINE 94012612
 PUBMED 8407904
 FEATURES
 source location/Qualifiers
 1..1901 /organism="Escherichia coli"
 /mol_type="genomic DNA"
 /db_xref="taxon:562"
 3..1522 /gene="appa"
 3..8 /gene="appa"
 16..108 /gene="appa"
 /note="ORF"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="AA72085.1"
 /db_xref="GI:145284"
 /translation="MMVLLMFVGLMCSLTVLVWLPRLKS"
 118..149 /gene="appa"
 133..138 /gene="appa"
 156..161 /gene="appa"
 /gene="appa"
 172..177 /gene="appa"
 /gene="appa"
 188..1486 /gene="appa"
 /standard_name="acid phosphatase"
 /codon_start=1
 /transl_table=11
 /product="periplasmic phosphoanhydride phosphohydrolase"
 /protein_id="AA72086.1"
 /db_xref="GI:145285"
 /translation="MKAILIPFLSLIPLPQSAFAQSEBELKLESVIVISRRGVNAP
 TKATQLMQDVTPEAMPPTWPKLGMIFPRGGELIAYIGHYGRQLVADGLLAKKCCPS

GOVAILADYDERTRKTRGFARAFAPDCAITVHNOADTSSPDPLFNPKTGVCOLDNA
NVTDAIISAGSISADFTGHRQTAFAREBRVINFPOSUCLKREKODSSCSLTOMLPS
ELKVSADNVSITGASVLSMTEIFLLQDAQMPEPMGRITDSHOMTLISLHNAOF
YLOQTPEVARSATPLDLITKALTPHPPOQAYGVLLPVSLEIACHDNLNLAJGG
ALELWMTLEGCPDNTPPGGELVFERMRRLSDSQWIOVSLVFQTLQOMRDKTPLSINT
PGEVYKLTLAGCEERNAQCMCSLAGFTQVNEARIIPACSL"
188..253
/gene="appa"
254..1483
/gene="appa"
/Product="periplasmic phosphoanhydride phosphohydrolyase"
299..301
/gene="appa"
/standard_name="H303A mutant"
/note="created by site directed mutagenesis"
/citation=[2]
299..301
/gene="appa"
/standard_name="R16A mutant"
/note="created by site directed mutagenesis"
/citation=[2]
299..301
/gene="appa"
/standard_name="R20A mutant"
/note="created by site directed mutagenesis"
/citation=[2]
299..301
/gene="appa"
/standard_name="R92A mutant"
/note="created by site directed mutagenesis"
/citation=[2]
302..304
/gene="appa"
/standard_name="H17N mutant"
/note="created by site directed mutagenesis"
/citation=[2]
1163..1165
/gene="appa"
/standard_name="D304A mutant"
/note="created by site directed mutagenesis"
/citation=[3]
/phenotype="Asp 304 changed to Ala"
/replace="gct"
1163..1165
/gene="appa"
/standard_name="D304E mutant"
/note="created by site directed mutagenesis"
/citation=[3]
/phenotype="Asp 304 changed to Glu"
/replace="gag"
1497..1522
/gene="appa"
/note="rho-independent terminator"

ORIGIN

Query Match 99.6%; Score 1293.2; DB 1; Length 1901;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1295; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATGAAAGCCATCTTAATCCATTTTATCTCTTCTGATTCCGTTAACCCGCAATCTGCA 60
DB 188 ATGAAAGCGATCTTAATCCATTTTATCTCTTCTGATTCCGTTAACCCGCAATCTGCA 247
QY 61 TTGCTCAGAGTGAAGCCGAGCTGAAGCTGGAAGTGTGGTATGTCAGTTCGTCAGTGT 120
DB 248 TTGCTCAGAGTGAAGCCGAGCTGAAGCTGGAAGTGTGGTATGTCAGTTCGTCAGTGT 307
QY 121 GTGGTGTCTCAACCAAGGCCAGCACTGATGAGAGTGTCAACCCAGAGCGATGAGCA 180
DB 308 GTGGTGTCTCAACCAAGGCCAGCACTGATGAGAGTGTCAACCCAGAGCGATGAGCA 367
QY 181 ACCTGGCCGTAATACTGGTGTGCTGACACCGCGGTGGTGAAGCTAATCGCCTATCTC 240

DB 368 ACTGGCCGTAATACTGGTGTGCTGACACCGCGNGGTGGTGAAGCTAATCGCCTATCTC 427
QY 241 GGACATTATCAACGCCGCGCTGGTAGCCGAGGATTTGCTGGGAAAAAGGCTGCCG 300
DB 428 GACATTATCAACGCCGCGCTGGTAGCCGAGGATTTGCTGGGAAAAAGGCTGCCG 487
QY 301 CAGTCTGTGTCAGTTCGCGATTATGTCGATGTCAGGACGCTACCCGTAACACGGCGAA 360
DB 488 CAGTCTGTGTCAGTTCGCGATTATGTCGATGTCAGGACGCTACCCGTAACACGGCGAA 547
QY 361 GCCTTCGCCGCCGCGCTGGACCTGATCTGTGCAATTAACCGTACATACCCAGCGATACG 420
DB 548 GCCTTCGCCGCCGCGCTGGACCTGATCTGTGCAATTAACCGTACATACCCAGCGATACG 607
QY 421 TCCAGTCCCGATCCGTTATTAATCTCTTAATAAACTGGCGTTTGCCAACTGATTAACGCG 480
DB 608 TCCAGTCCCGATCCGTTATTAATCTCTTAATAAACTGGCGTTTGCCAACTGATTAACGCG 667
QY 481 AACGTGACTGACGCGATCTCAGCAGGCGAGAGGTCATTTGCTGACTTTACCGGCGAT 540
DB 668 AACGTGACTGACGCGATCTCAGCAGGCGAGAGGTCATTTGCTGACTTTACCGGCGAT 727
QY 541 CGGCAAAAGCGGCTTTCGGAAGTGGAACTGGAACGGGTGCTTAATTTCCGCAATCAACTGTGC 600
DB 728 CGGCAAAAGCGGCTTTCGGAAGTGGAACTGGAACGGGTGCTTAATTTCCGCAATCAACTGTGC 787
QY 601 CTTAAACGTGAAGAAACAGAGCAAGAACTGTTCAATTAACGAGCATTAACGAGCATC 660
DB 788 CTTAAACGTGAAGAAACAGAGCAAGAACTGTTCAATTAACGAGCATTAACGAGCATC 847
QY 661 AAGGTGAGCGCGCAATGTCTCAATTAACGGGTGCTTAACGCTTCGCAATGCTGACG 720
DB 848 AAGGTGAGCGCGCAATGTCTCAATTAACGGGTGCTTAACGCTTCGCAATGCTGACG 907
QY 721 GAGATATTTCTCTCTGCAACAGCAAGGAATGCCGAGCCGGGTGGGAAAGATCAC 780
DB 908 GAGATATTTCTCTCTGCAACAGCAAGGAATGCCGAGCCGGGTGGGAAAGATCAC 967
QY 781 GATTCAACACAGTGAACACCTTGCTAAGTTGCAACGCGCAATTTATTTGCTACAA 840
DB 968 GATTCAACACAGTGAACACCTTGCTAAGTTGCAACGCGCAATTTATTTGCTACAA 1027
QY 841 CGCAGCGCCAGAGTTGCCCGCAGCCGCGCACCCCGTTATTAGATTGATCAAGCACGCG 900
DB 1028 CGCAGCGCCAGAGTTGCCCGCAGCCGCGCACCCCGTTATTAGATTGATCAAGCACGCG 1087
QY 901 TTGAGCGCCCATCCACCCGCAAAAACAGCGGTAATGAGTGAACATTAACCACTCAGTCTG 960
DB 1088 TTGAGCGCCCATCCACCCGCAAAAACAGCGGTAATGAGTGAACATTAACCACTCAGTCTG 1147
QY 961 TTATGCGCGGACAGCATCTAATCTGCAAAATCTCGCGGCGCACTGAGCTCAACTGG 1020
DB 1148 TTATGCGCGGACAGCATCTAATCTGCAAAATCTCGCGGCGCACTGAGCTCAACTGG 1207
QY 1021 ACGCTTCCGCTGACGCGGATTAACGCGCCGCAAGTGTGAACTGTGTTGAAGCTGG 1080
DB 1208 ACGCTTCCGCTGACGCGGATTAACGCGCCGCAAGTGTGAACTGTGTTGAAGCTGG 1267
QY 1081 CGTCCGCTAAGAGATTAACGCGAGTGAATTAAGGTTTGTGCTGCTCCAGACTTACAG 1140
DB 1268 CGTCCGCTAAGAGATTAACGCGAGTGAATTAAGGTTTGTGCTGCTCCAGACTTACAG 1327
QY 1141 CAGATCGTGAATAAAGCGCGCTGTCATTAATAAGCCGCGCGAGAGGTGAAGCTGACC 1200
DB 1328 CAGATCGTGAATAAAGCGCGCTGTCATTAATAAGCCGCGCGAGAGGTGAAGCTGACC 1387
QY 1201 CTGGCAGATGTGAAGAGGAAATGCGCAGGCAATGTGTTGTTGGCAGGTTTACGCA 1260
DB 1388 CTGGCAGATGTGAAGAGGAAATGCGCAGGCAATGTGTTGTTGGCAGGTTTACGCA 1447
QY 1261 ATGTGAATGAAGACAGCATACCCGTTGCAATTTGTAA 1299
DB 1448 ATGTGAATGAAGACAGCATACCCGTTGCAATTTGTAA 1486

RESULT 10
 AX356572 1901 bp DNA linear PAT 06-FEB-2002
 LOCUS Sequence 7 from Patent WO0190333.
 DEFINITION AX356572
 ACCESSION AX356572 GI:18621059
 VERSION AX356572.1
 KEYWORDS
 SOURCE
 ORGANISM
 Escherichia coli
 Escherichia coli
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Escherichia.
 1
 Short, J.M., Kretz, K.A., Gray, K.A., Barton, N.R., Garrett, J.B. and
 O'Donoghue, E.
 Recombinant bacterial phylases and uses thereof
 Patent: WO 0190333-A 729-NOV-2001;
 DIVERSA CORPORATION (US)
 Location/Qualifiers
 1. 1901
 /organism="Escherichia coli"
 /mol_type="unassigned DNA"
 /db_xref="taxon:562"

ORIGIN

Query Match 99.6%; Score 1293.2; DB 6; Length 1901;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 1295; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

1 ATGAAGCCATCTTAAATCCCATTTTATCTCTTCTGATTCCTGTTAACCCCGCATCTGCA 60
 188 ATAAAGCGATCTTAAATCCCATTTTATCTCTTCTGATTCCTGTTAACCCCGCATCTGCA 247
 61 TTGGCTCAGATGAGCGCGAGCTGAGCTGAGTGAAGTGTGATGTCAGTCGTCATGCT 120
 248 TTGGCTCAGATGAGCGCGAGCTGAGCTGAGTGAAGTGTGATGTCAGTCGTCATGCT 307
 121 GTGCGTCTCAACCAAGGCGCACTGATGCAAGATGTCACCCGAGCGCATGCGCA 180
 308 GTGCGTCTCAACCAAGGCGCACTGATGCAAGATGTCACCCGAGCGCATGCGCA 367
 181 ACCTGGCCGGTAAACTGGGTTGCTGACACCGCGCGGTGTGAGCTATTCGCTTATCTC 240
 368 ACCTGGCCGGTAAACTGGGTTGCTGACACCGCGCGGTGTGAGCTATTCGCTTATCTC 427
 241 GGAATTATCAACGCGCGCTGTGTCGAGCGGATTTGCTGCGGAAAAAGGCGTGCCTG 300
 428 GGAATTATCAACGCGCGCTGTGTCGAGCGGATTTGCTGCGGAAAAAGGCGTGCCTG 487
 301 CAGTCTGTGATGTCGAGTATTTCTGATGTCAGACGATGTCACCGTAAACAGCGGAA 360
 488 CAGTCTGTGATGTCGAGTATTTCTGATGTCAGACGATGTCACCGTAAACAGCGGAA 547
 361 GCCTTGCCTGCGCGGCTGCGCACTGATGTCAGTAAACCGTATACCCAGCGCATACG 420
 548 GCCTTGCCTGCGCGGCTGCGCACTGATGTCAGTAAACCGTATACCCAGCGCATACG 607
 421 TTCCGTTCCCATTCGTTATTTATCTCTTAAACTGCGCGTTTCCCACTGGATTAAGCG 480
 608 TTCCGTTCCCATTCGTTATTTATCTCTTAAACTGCGCGTTTCCCACTGGATTAAGCG 667
 481 AACGTGATGATGTCGATCTTCAAGAGGCGAGGATGTCATTTGCTGATTTTACCGGAT 540
 668 AACGTGATGATGTCGATCTTCAAGAGGCGAGGATGTCATTTGCTGATTTTACCGGAT 727
 541 CGGCAAAACGCGGTTTCCGGAATCTGAAACGCGGCTTAAATTTCCGCAATCAACTTGTGC 600
 728 CGGCAAAACGCGGTTTCCGGAATCTGAAACGCGGCTTAAATTTCCGCAATCAACTTGTGC 787
 601 CTTAAAGCTGAGAAACGAGAAAGCTGTTCAATTAACGAGGATTAACATCGGAATCTC 660
 788 CTTAAAGCTGAGAAACGAGAAAGCTGTTCAATTAACGAGGATTAACATCGGAATCTC 847

661 AAGGTGAGCGCGGACAAATGTCATTAACCGGCGGTAAAGCTGCGCATCAATGCTGACG 720
 848 AAGGTGAGCGCGGACAAATGTCATTAACCGGCGGTAAAGCTGCGCATCAATGCTGACG 907
 721 GAGATATTTCTCTGCAACAAAGGAGATGCGGAGCGCGGCTGGGAAAGATCACC 780
 908 GAGATATTTCTCTGCAACAAAGGAGATGCGGAGCGCGGCTGGGAAAGATCACC 967
 781 GATTACACACCGATGAAACACTTGTCTAAGTTTGCATTAACGCGCAATTTATTTGCTACAA 840
 968 GATTACACACCGATGAAACACTTGTCTAAGTTTGCATTAACGCGCAATTTATTTGCTACAA 1027
 841 CGGACCGCGAGGTTGCG 900
 1028 CGGACCGCGAGGTTGCG 1087
 901 TTGACCGCGCATTCACCGCAAAACAGGCGTATGTTGATCAATTAACCACTTCAGTCTG 960
 1088 TTGACCGCGCATTCACCGCAAAACAGGCGTATGTTGATCAATTAACCACTTCAGTCTG 1147
 961 TTTATGCGCGGACACGATATCTAATCTTGGCAATCTTGGCGGCGCACTGAGCTCAACTGG 1020
 1148 TTTATGCGCGGACACGATATCTAATCTTGGCAATCTTGGCGGCGCACTGAGCTCAACTGG 1207
 1021 ACCGTTCCCGGTACCGCGGATTAACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1080
 1208 ACCGTTCCCGGTACCGCGGATTAACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1267
 1081 CGTGGCTAAGCGATTAACAGCGCGGATTAAGGTTTGGCTGCTTCCGACCTTTACAG 1140
 1268 CGTGGCTAAGCGATTAACAGCGCGGATTAAGGTTTGGCTGCTTCCGACCTTTACAG 1327
 1141 CAGATCGGTATTAACCGCGCTGTCTAATTAATACCGCGCGCGCGCGCGCGCGCGCG 1200
 1328 CAGATCGGTATTAACCGCGCTGTCTAATTAATACCGCGCGCGCGCGCGCGCGCGCG 1387
 1201 CTGGCAGATGTTAAAGGCAATGCGGAGGCGGATGTTGCTGAGAGTTTACGCA 1260
 1388 CTGGCAGATGTTAAAGGCAATGCGGAGGCGGATGTTGCTGAGAGTTTACGCA 1447
 1261 ATCGTAATGAACGACGATACCGGCTTGCAGTTTGA 1299
 1448 ATCGTAATGAACGACGATACCGGCTTGCAGTTTGA 1486

RESULT 11
 ECOACPHB 1901 bp DNA linear BCT 04-APR-2002
 LOCUS Escherichia coli periplasmic phosphoanhydride phosphohydrolase
 DEFINITION (appp) gene, complete cds.
 ACCESSION 103371
 VERSION 103371.1 GI:145178
 KEYWORDS acid phosphatase; periplasmic phosphoanhydride phosphohydrolase.
 SOURCE Escherichia coli
 ORGANISM Escherichia coli
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Escherichia.
 1 (sites)
 Dassa, J., Marck, C. and Boquet, P.L.
 The complete nucleotide sequence of the Escherichia coli gene appp
 reveals significant homology between pH 2.5 acid phosphatase and
 glucose-1-phosphatase
 J. Bacteriol. 172 (9), 5497-5500 (1990)
 JOURNAL MEDLINE 90368616
 PUBMED 2168385
 2 (bases 1 to 1901)
 Oesterlin, K., Harms, E.H., Stevis, P.E., Kuciel, R., Zhou, M.M. and Van
 Overbeek, R.L.
 Overexpression, site-directed mutagenesis, and mechanism of
 Escherichia coli acid phosphatase
 J. Biol. Chem. 267 (32), 22830-22836 (1992)
 JOURNAL MEDLINE 93054596

PUBMED 1429631 Location/Qualifiers
FEATURES
source 1..1901
/organism="Escherichia coli"
/mol_type="genomic DNA"
/db_xref="taxon:562"
188..1486
/gene="appa"
188..1486
/gene="appa"
/standard_name="acid phosphatase"
/citation=[1]
/codon_start=1
/transl_table=11
/product="periplasmic phosphoanhydride phosphohydrolyase"
/protein_id="AAA0003.1"
/db_xref="GI:145179"
/translation="MKAILIPLSILIPLPPOSAPASPELKESVIVSRNGVRAP
TKATQMDVTPDAMPVPMVKGMTPRGGELIAYIGHYORORLADGLAKKGRPS
GQVATIDYDERTRKTEAFNAGLAPDCATITHTOADTSSPPLPNPLKTVCOLDNA
NVTDAILSHAGSIADFTGHRQTAFLERLVRVNPQSNCLREKODSCSLTQALPS
ELKVSADNVSLTGAIVSLASMLTEIFLQQAQMPBPGWRITDSHQMTLLSLNAQF
YLQRTPEVARSRATPLIDLITKALTPHPQQAQGVLLPTSLFIAQIDTLANIG
ALEIMWTLPGQDNPPEGELVEERMRRLSDNSOMIYVSLVFTQLQOMDKPLSLNT
PGEVKTITLAGEBERNAQMGSLAGFTQIVNEARIPAGSL"
302..304
/gene="appa"
/citation=[1]
/phenotype="H17N mutant"
ORIGIN
Query Match 99.4%; Score 1291.6; DB 1; Length 1901;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1294; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
1 ATGAAGCCATCTTAATCCATTTTATCTCTGATTCGGTTAACCCGCAATCTGA 60
188 ATGAAGCCATCTTAATCCATTTTATCTCTGATTCGGTTAACCCGCAATCTGA 247
61 TTGCTCAGAGTGGCCGAGCTGAAGCTGGAAGTGTGATTTGCTAGTCTCATGGT 120
248 TTGCTCAGAGTGGCCGAGCTGAAGCTGGAAGTGTGATTTGCTAGTCTCATGGT 307
121 GTGCTGCTCAACCAAGGCCAAGCACTGATGACAGATGTCACCCCAAGCGATGCCA 180
308 GTGCTGCTCAACCAAGGCCAAGCACTGATGACAGATGTCACCCCAAGCGATGCCA 367
181 ACCTGGCCGGTAAACCTGGGTTGAGCTGACACCGCGGCTGATGATTCGCTATCTC 240
368 ACCTGGCCGGTAAACCTGGGTTGAGCTGACACCGCGGCTGATGATTCGCTATCTC 427
241 GACATTACCAACGCCAGCGTGTGTAAGCCAGGATTTGCTGCGAAGAAAGGCTGCCG 300
428 GACATTACCAACGCCAGCGTGTGTAAGCCAGGATTTGCTGCGAAGAAAGGCTGCCG 487
301 CAGTCTGTGATGATTCGATTTATGCTGATGATGATGATGATGATGATGATGATGAT 360
488 CAGTCTGTGATGATTCGATTTATGCTGATGATGATGATGATGATGATGATGATGAT 547
361 GCTTGGCCGCGGCGTGGACCTGATGTCATTAACCGTAAATCCAGGAGATGAC 420
548 GCTTGGCCGCGGCGTGGACCTGATGTCATTAACCGTAAATCCAGGAGATGAC 607
421 TCCAGTCCGATCCGTTATTTAATCTCTTAAACCTGGCTTTCCCACTGATTAAGCG 480
608 TCCAGTCCGATCCGTTATTTAATCTCTTAAACCTGGCTTTCCCACTGATTAAGCG 667
481 AACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540
668 AACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 727
541 CGGCAACGGCGTTTCCGGAACCTGGAACGGGTGCTTAATTTTCCGCAATCAAACTTGTG 600

DB 728 CGGCAACGGCGTTTCCGGAACCTGGAACGGGTGCTTAATTTTCCGCAATCAAACTTGTG 787
QY 601 CTTAAACGTGAGAAACAGAGCAAGCTTTCATTAACGAGGATTAACCATCGGAATC 660
DB 788 CTTAAACGTGAGAAACAGAGCAAGCTTTCATTAACGAGGATTAACCATCGGAATC 847
QY 661 AAGGTGAGCGCCGCAATATGCTCATTAACCGGTGCGGTAAACCTCGCATCAATGCTGACG 720
DB 848 AAGGTGAGCGCCGCAATATGCTCATTAACCGGTGCGGTGAGCTCGCATCAATGCTGACG 907
QY 721 GAGATATTTCTCTGCAACCAAGCAAGGAAATGCCCGAGCCGGGTGGGAAAGATCAAC 780
DB 908 GAGATATTTCTCTGCAACCAAGCAAGGAAATGCCCGAGCCGGGTGGGAAAGATCAAC 967
QY 781 GATTACACCGATGGAACACCTTGCTAAGTTTGATTAACGGCAATTTATTTGCTACAA 840
DB 968 GATTACACCGATGGAACACCTTGCTAAGTTTGATTAACGGCAATTTATTTGCTACAA 1027
QY 841 CGCACGCGAGAGTTTCCGCGGCGCGCCACCCCGTTATTAGATTGATCAAGACAGCG 900
DB 1028 CGCACGCGAGAGTTTCCGCGGCGCGCCACCCCGTTATTAGATTGATCAAGACAGCG 1087
QY 901 TTGACGCGCCCATCCACCGCAAAAAAGCGCTATGCTGATTAACCATTCCTGCTG 960
DB 1088 TTGACGCGCCCATCCACCGCAAAAAAGCGCTATGCTGATTAACCATTCCTGCTG 1147
QY 961 TTTATCGCGGAGACGATTAATCTGCAATATCTGCGCGCGCACTGAGACTGAACCTG 1020
DB 1148 TTTATCGCGGAGACGATTAATCTGCAATATCTGCGCGCGCACTGAGACTGAACCTG 1207
QY 1021 ACGCTTCCCGGTGACCGGATTAACACCGCGCGAGTGTGAATCTGTGTAACGCTG 1080
DB 1208 ACGCTTCCCGGTGACCGGATTAACACCGCGCGAGTGTGAATCTGTGTAACGCTG 1267
QY 1081 CGTGGCTTAAGCATTAACAGCGAGTGAATTAAGTTTGGCTGCTTCCAGACTTTACG 1140
DB 1268 CGTGGCTTAAGCATTAACAGCGAGTGAATTAAGTTTGGCTGCTTCCAGACTTTACG 1327
QY 1141 CAGATGCGGATTAACAGCGCGCTGCTATTAAATACCGCGCGAGAGTGAATCTGAC 1200
DB 1328 CAGATGCGGATTAACAGCGCGCTGCTATTAAATACCGCGCGAGAGTGAATCTGAC 1387
QY 1201 CTGGCAGATGTAAGAGCGAAATGCGAGGCGATGTTGTTGGCAGGTTTACGAA 1260
DB 1388 CTGGCAGATGTAAGAGCGAAATGCGAGGCGATGTTGTTGGCAGGTTTACGAA 1447
QY 1261 ATCGTGAATGAAGCAGCATACCGCTTGCATTTGTAA 1299
DB 1448 ATCGTGAATGAAGCAGCATACCGCTTGCATTTGTAA 1486
RESULT 12
AX356570
LOCUS AX356570 1901 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 5 from Patent WO0190333.
ACCESSION AX356570
VERSION AX356570.1 GI:18621057
KEYWORDS
SOURCE
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE
1 Short, J.M., Kretz, K.A., Gray, K.A., Barton, N.R., Garrett, J.B. and
AUTHORS O'Donoghue, B.
TITLE Recombinant bacterial phytases and uses thereof
JOURNAL Patent: WO 0190333-A 5 29-NOV-2001;
DIVERSA CORPORATION (US)
FEATURES
source 1..1901
/organism="Escherichia coli"
/mol_type="unassigned DNA"
/db_xref="taxon:562"

ORIGIN

Query Match 99.4%; Score 1291.6; DB 6; Length 1901;
 Best Local Similarity 99.6%; Pred. No. 0;
 Matches 1294; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

1 ATGAAAGCCATCTTAATCCCATTTTATCTCTTCTGATTCCTGTAACCCCGCAATGTGCA 60
 188 ATGAAAGCCATCTTAATCCCATTTTATCTCTTCTGATTCCTGTAACCCCGCAATGTGCA 247
 61 TTGCGCTCAGATGAGCCGAGCTGAAAGTGAAGTGGTATGTCAGTCTCATGTGT 120
 248 TTGCGCTCAGATGAGCCGAGCTGAAAGTGAAGTGGTATGTCAGTCTCATGTGT 307
 121 GTGCGTCTCAACCAAGGCAACGCAATGATGCAAGATGTCAACCCAGACGATGACCA 180
 308 GTGCGTCTCAACCAAGGCAACGCAATGATGCAAGATGTCAACCCAGACGATGACCA 367
 181 ACCTGGCCGATGAAATCTGGGTGGTGAACACCGCGGGTGGTGAAGTAAATCGCTATCTC 240
 368 ACCTGGCCGATGAAATCTGGGTGGTGAACACCGCGGGTGGTGAAGTAAATCGCTATCTC 427
 241 GGACATTACCAACGCGACGCTGTGTAGCCGACGATGCTGGCGAAGGAGGCTGCCG 300
 428 GGACATTACCAACGCGACGCTGTGTAGCCGACGATGCTGGCGAAGGAGGCTGCCG 487
 301 CAGTCTGTGATGATGCGATTTATGCTGATGTCAGAGCGTACCCGTAACAGAGCGAA 360
 488 CAGTCTGTGATGATGCGATTTATGCTGATGTCAGAGCGTACCCGTAACAGAGCGAA 547
 361 GCGTTGGCGCGCGGGCTGGCACTGATGTCATTAACCTTAATACCCAGAGGATACG 420
 548 GCGTTGGCGCGCGGGCTGGCACTGATGTCATTAACCTTAATACCCAGAGGATACG 607
 421 TCCAGTCCGATCGTTATTTAATCCCTAATAACCTGGCTGGCACTGATTAACGCG 480
 608 TCCAGTCCGATCGTTATTTAATCCCTAATAACCTGGCTGGCACTGATTAACGCG 667
 481 AACGTGATGATGCGATCTCAGACAGGCGAGAGGTCATTTGCTGACTTTACCGGCGAT 540
 668 AACGTGATGATGCGATCTCAGACAGGCGAGAGGTCATTTGCTGACTTTACCGGCGAT 727
 541 CCGCAAAACGCGCTTTGCGCAACTGGAACGCGGTCTTAATTTTCCGCAATCAACTTGTGC 600
 728 CCGCAAAACGCGCTTTGCGCAACTGGAACGCGGTCTTAATTTTCCGCAATCAACTTGTGC 787
 601 CTTAACGATGAAGCAAGCAAGAGCTGTTCAATTAAGGATTAACCATCGGAATCTC 660
 788 CTTAACGATGAAGCAAGCAAGAGCTGTTCAATTAAGGATTAACCATCGGAATCTC 847
 661 AAGGTGAGCCGCAATGCTCTCAATTAACGCGTGGTAAAGCTTCGATCAATGCTGACG 720
 848 AAGGTGAGCCGCAATGCTCTCAATTAACGCGTGGTAAAGCTTCGATCAATGCTGACG 907
 721 GAGATATTTCTCTGCAACAGCAACAGGAAATGCCGAGCCGGGTGGGGAAGATCAAC 780
 908 GAGATATTTCTCTGCAACAGCAACAGGAAATGCCGAGCCGGGTGGGGAAGATCAAC 967
 781 GATTCAACAGAGTGAACACCTTGCTAAGTTTGAATAGCGCAATTTTATTTGCTCAAA 840
 968 GATTCAACAGAGTGAACACCTTGCTAAGTTTGAATAGCGCAATTTTATTTGCTCAAA 1027
 841 CGACGCGCAGAGTGGCCGCGACGCGCGCAACCCGTTATTTGAATTTGATCAAGACGCG 900
 1028 CGACGCGCAGAGTGGCCGCGACGCGCGCAACCCGTTATTTGAATTTGATCAAGACGCG 1087
 901 TTGAACGCGCCATCAACGCAAAACAGGCGTATGCTGATTAACCATCTTCACTGTCTG 960
 1088 TTGAACGCGCCATCAACGCAAAACAGGCGTATGCTGATTAACCATCTTCACTGTCTG 1147
 961 TTATATGCGCGGACGATTAATCTGGCAAAATCTCGGCGGCGGATCGAGCTCAACTGG 1020
 1148 TTATATGCGCGGACGATTAATCTGGCAAAATCTCGGCGGCGGATCGAGCTCAACTGG 1207

QY 1021 AGCTTCCCGGTGAGCCGATTAACGCGCGAGGTGGAATGCTGTTGAACGCTGG 1080
 DB 1208 AGCTTCCCGGTGAGCCGATTAACGCGCGAGGTGGAATGCTGTTGAACGCTGG 1267
 QY 1081 CGTGGCTAAGGATTAACGCGAGTGAATGAGTTTCGCTGCTTCCAGACTTTACAG 1140
 DB 1268 CGTGGCTAAGGATTAACGCGAGTGAATGAGTTTCGCTGCTTCCAGACTTTACAG 1327
 QY 1141 CAGATGCTGATTAACGCGCTGCTTATTAATGCGCCGCCCGGAGAGGTGAACCTGACC 1200
 DB 1328 CAGATGCTGATTAACGCGCTGCTTATTAATGCGCCGCCCGGAGAGGTGAACCTGACC 1387
 QY 1201 CTGGCAGATGTAAGAGGAAATGGCGAGGAGATGTTTCGTTGACAGTTTACGAA 1260
 DB 1388 CTGGCAGATGTAAGAGGAAATGGCGAGGAGATGTTTCGTTGACAGTTTACGAA 1447
 QY 1261 ATCGTGAATGAAGACGATTAACCGCTTCAGTTGTAA 1299
 DB 1448 ATCGTGAATGAAGACGATTAACCGCTTCAGTTGTAA 1486

RESULT 13
 AX356574 1901 bp DNA linear PAT 06-FEB-2002
 LOCUS AX356574
 DEFINITION Sequence 9 from Patent WO0190333.
 ACCESSION AX356574
 VERSION AX356574.1 GI:18621060
 KEYWORDS
 SOURCE
 ORGANISM
 Escherichia coli
 Escherichia coli
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Escherichia.

REFERENCE
 AUTHORS Short, J.M., Kretz, K.A., Gray, K.A., Barton, N.R., Garrett, J.B. and O'Donoghue, E.
 TITLE Recombinant bacterial phylases and uses thereof
 JOURNAL Patent: WO 0190333-A 9 29-NOV-2001;
 DIVERSA CORPORATION (US)
 FEATURES
 source Location/Qualifiers
 1. 1901
 /organism="Escherichia coli"
 /mol_type="unassigned DNA"
 /db_xref="taxon:562"

ORIGIN
 Query Match 99.4%; Score 1291.6; DB 6; Length 1901;
 Best Local Similarity 99.6%; Pred. No. 0;
 Matches 1294; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ATGAAAGCCATCTTAATCCCATTTTATCTCTTCTGATTCCTGTAACCCCGCAATGTGCA 60
 DB 188 ATGAAAGCCATCTTAATCCCATTTTATCTCTTCTGATTCCTGTAACCCCGCAATGTGCA 247
 61 TTGCGCTCAGATGAGCCGAGCTGAAAGTGAAGTGGTATGTCAGTCTCATGTGT 120
 248 TTGCGCTCAGATGAGCCGAGCTGAAAGTGAAGTGGTATGTCAGTCTCATGTGT 307
 121 GTGCGTCTCAACCAAGGCAACGCAATGATGCAAGATGTCAACCCAGACGATGACCA 180
 308 GTGCGTCTCAACCAAGGCAACGCAATGATGCAAGATGTCAACCCAGACGATGACCA 367
 181 ACCTGGCCGATGAAATCTGGGTGGTGAACACCGCGGGTGGTGAAGTAAATCGCTATCTC 240
 368 ACCTGGCCGATGAAATCTGGGTGGTGAACACCGCGGGTGGTGAAGTAAATCGCTATCTC 427
 241 GGACATTACCAACGCGACGCTGTGTAGCCGACGATGCTGGCGAAGGAGGCTGCCG 300
 428 GGACATTACCAACGCGACGCTGTGTAGCCGACGATGCTGGCGAAGGAGGCTGCCG 487
 301 CAGTCTGTGATGATGCGATTTATGCTGATGTCAGAGCGTACCCGTAACAGAGCGAA 360
 488 CAGTCTGTGATGATGCGATTTATGCTGATGTCAGAGCGTACCCGTAACAGAGCGAA 547


```

QY 361 GCCTTCGCGCGGCTGGACCTGACTGTCAATTAACCGATACCCAGGAGATACG 420
DB 548 GCTTCGCGCGGCTGGACCTGACTGTCAATTAACCGATACCCAGGAGATACG 607
QY 421 TCCAGTCCCGATCCGTTATTTTATCTCTTAAAACTGGCGTTTGGCACTGATTAACGCG 480
DB 608 TCCAGTCCCGATCCGTTATTTTATCTCTTAAAACTGGCGTTTGGCACTGATTAACGCG 667
QY 481 AACGTAATGAGCGATCTCTGACGAGGAGGAGGATCAATTGCTGACTTTAACGCGGAT 540
DB 668 AACGTAATGAGCGATCTCTGACGAGGAGGAGGATCAATTGCTGACTTTAACGCGGAT 727
QY 541 CGGCAAAACGCGTTTCCGCACTGGAACGCGGTCTTAAATTTTCCGCAATCAAACTTGTGC 600
DB 728 CGGCAAAACGCGTTTCCGCACTGGAACGCGGTCTTAAATTTTCCGCAATCAAACTTGTGC 787
QY 601 CTTAAACGTAAGAAACAGGAGAAAGCTGTTCATTACGAGGCACTTACCGGAACTC 660
DB 788 CTTAAACGTAAGAAACAGGAGAAAGCTGTTCATTACGAGGCACTTACCGGAACTC 847
QY 661 AAGGTAGGCGCGAACAATGCTCAATTACCGGTGCGTAAAGCTGCAATCAATGCTGACG 720
DB 848 AAGGTAGGCGCGAACAATGCTCAATTACCGGTGCGTAAAGCTGCAATCAATGCTGACG 907
QY 721 GAGATATTTCTCTGCAACAAGCAAGGAATGCGGAGCCGCGGTGGGAAAGATCAAC 780
DB 908 GAGATATTTCTCTGCAACAAGCAAGGAATGCGGAGCCGCGGTGGGAAAGATCAAC 967
QY 781 GATTCAACCAAGTGAACCTTGTGAATTTGCAATTAACGCGCAATTTATTTGCTACAA 840
DB 968 GATTCAACCAAGTGAACCTTGTGAATTTGCAATTAACGCGCAATTTATTTGCTACAA 1027
QY 841 CGCAGCCAGAGGTTCCGCGAGCCGCGCAACCCGTTATTGATTTGATCAAGACGCG 900
DB 1028 CGCAGCCAGAGGTTCCGCGAGCCGCGCAACCCGTTATTGATTTGATCAAGACGCG 1087
QY 901 TTGACGCGCCATTCACCGCAAAACAGGCGTATGCTGATCACTTACCGCACTTCAAGTCG 960
DB 1088 TTGACGCGCCATTCACCGCAAAACAGGCGTATGCTGATCACTTACCGCACTTCAAGTCG 1147
QY 961 TTGATGCGCGGAGCAAGTACTTAATCTGAGCAATCTCGGCGGCGACACTGAGCTCAACTG 1020
DB 1148 TTGATGCGCGGAGCAAGTACTTAATCTGAGCAATCTCGGCGGCGACACTGAGCTCAACTG 1207
QY 1021 ACGCTTCCCGGTCAAGCGGATTAACGCGCGCAAGTGTGAATCTGTTTGAACGCTGCG 1080
DB 1208 ACGCTTCCCGGTCAAGCGGATTAACGCGCGCAAGTGTGAATCTGTTTGAACGCTGCG 1267
QY 1081 CGTGGCTGAAGCGATTAACGCGAGTGAATCAAGTTTCCGTTGCTTCCAGACTTTACAG 1140
DB 1268 CGTGGCTGAAGCGATTAACGCGAGTGAATCAAGTTTCCGTTGCTTCCAGACTTTACAG 1327
QY 1141 CAGATGCGGTAAAGAAACGCGCTGTCTTAATATACGCGCGCGAGAGGTTGAAACTGAC 1200
DB 1328 CAGATGCGGTAAAGAAACGCGCTGTCTTAATATACGCGCGCGAGAGGTTGAAACTGAC 1387
QY 1201 CTGGCAGGATGTGAAGAGCGAATGCGAGGCGCATGTGTCTGTGGAGAGTTTAAAGCA 1260
DB 1388 CTGGCAGGATGTGAAGAGCGAATGCGAGGCGCATGTGTCTGTGGAGAGTTTAAAGCA 1447
QY 1261 ATCGTGAATGAAGACGATACCGGCTTGCAGTTTGTAA 1299
DB 1448 ATCGTGAATGAAGACGATACCGGCTTGCAGTTTGTAA 1486

```

```

RESULT 14
AF537219 1315 bp DNA linear BCT 29-SEP-2002
LOCUS Escherichia coli AppA (appA) gene, complete cds.
DEFINITION AF537219
ACCESSION AF537219
VERSION AF537219.1 GI:23344056
KEYWORDS

```

```

SOURCE Escherichia coli
ORGANISM Escherichia coli
REFERENCE Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
AUTHORS Enterobacteriaceae; Escherichia.
TITLE 1 (bases 1 to 1315)
JOURNAL Zhang, Z., He, J., Yao, B., Zhou, Y., Chen, Y. and Yi, Y.
AUTHORS Production of phytase and acid phosphatase by use of
TITLE silkworm-bioreactor
JOURNAL Patent: China (01127288.0) -A 29-SEP-2001;
AUTHORS 2 (bases 1 to 1315)
REFERENCE Chen, Y., Zhu, Z., Zhang, Z. and He, J.
TITLE Cloning and overexpression of phytase gene appA from Escherichia
JOURNAL coli
REFERENCE Unpublished
AUTHORS 3 (bases 1 to 1315)
JOURNAL Chen, Y., Zhu, Z., Zhang, Z. and He, J.
TITLE Direct Submission
JOURNAL Submitted (10-AUG-2002) Key Laboratory of Silkworm Biotechnology,
Ministry of Agriculture, Sericultural Research Institute, Chinese
Academy of Agricultural Sciences, Sibaidu, Zhenjiang, Jiangsu
Province 212018, China
FEATURES
source
location/Qualifiers
1..1315
/organism="Escherichia coli"
/mol_type="genomic DNA"
/db_xref="taxon:562"
10..1308
/gene="appA"
10..1308
/gene="appA"
/note="phytase; phosphatase"
/codon_start=1
/transl_table=1
/product="appA"
/protein_id="AAN2834.1"
/db_xref="GI:23344057"
/translation="MKATLIFPLSLIPLPFQSAFQSSPEIKLESVIVSRHGVRA
TKATQMDVDPDAMPVPLKGLVPRGSLIVLGHYRQRLVADGLLAKKCPDS
GOVAIADVDERTKTGEFAAGLAPDAITVHQADITSPLFNPDKTGVCLDNA
NVTDAILFRAGSIADEFGRQTAFLERLELVNPQSNLCLKREKQESCSLTGALPS
ELKVSADNVSITGAVSLASMLTKI FLQQAQMPGPGMGRITDSHOMWNTLISLNAOF
YLQRTPEVARSRAATPLIDILKTALITPHRPOKAYGVTLPSVLEFIAGHDYNLNLGG
ALELWTLPGPDNPGRGLVPERMRRLSPNSQWIOVSLYFQTLQQRDKTPSLNT
PPEVKLTLACGERNAQMCSLAGFTQIVNKAIPACSL"
ORIGIN
Query Match 99.4%; Score 1291; DB 1; Length 1315;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1294; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 ATGAAAGCATTTATCCCATTTTATCTCTTGTGATTCGTTAACCCCGCAATCTGCA 60
DB 10 ATGAAAGCATTTATCCCATTTTATCTCTTGTGATTCGTTAACCCCGCAATCTGCA 69
QY 61 TTGCGTCAAGTGAAGCGGAGCTGAAGCTGGAAGATGTGGTGAATGTCAGTGCATGAT 120
DB 70 TTGCGTCAAGTGAAGCGGAGCTGAAGCTGGAAGATGTGGTGAATGTCAGTGCATGAT 129
QY 121 GTGCGGTCTCAACCAAGGCGACGCAACTGATGAGATGTCAACCCAGAGCGCATGGCCA 180
DB 130 GTGCGGTCTCAACCAAGGCGACGCAACTGATGAGATGTCAACCCAGAGCGCATGGCCA 189
QY 181 ACCTGCGCGGTAAACTGGGTGGTGTGACACCGCGCGGTGTGAGTGAATATGCGCTATCTC 240
DB 190 ACCTGCGCGGTAAACTGGGTGGTGTGACACCGCGCGGTGTGAGTGAATATGCGCTATCTC 249
QY 241 GGAATTAACCAAGCGGAGCTGTGTAAGCGGACGATGCTGCGGCAAAAGGCTGCGCG 300
DB 250 GGAATTAACCAAGCGGAGCTGTGTAAGCGGACGATGCTGCGGCAAAAGGCTGCGCG 309
QY 301 CAGTGTGTCAGGTCCGCAATTAATTGCTGATGTGACGAGCGTACCCGTAACAGGCGAA 360
DB 310 CAGTGTGTCAGGTCCGCAATTAATTGCTGATGTGACGAGCGTACCCGTAACAGGCGAA 369

```

QY 361 GCTTGGCGGCGGCTGACCTGACCTGCAATACCGTACATACCGAGGACATACG 420
 DB 370 GCTTGGCGGCGGCGGCTGACCTGACCTGCAATACCGTACATACCGAGGACATACG 429
 QY 421 TCCAGTCCGATCCGATATTTAATCTCTAATAAAGTGGGTTGGCAACGATACGCG 480
 DB 430 TCCAGTCCGATCCGATATTTAATCTCTAATAAAGTGGGTTGGCAACGATACGCG 489
 QY 481 AACGTACTGACGCGATCTCTACAGCAGGAGGAGGATGCTCAATGCTTACCGGCGAT 540
 DB 490 AACGTACTGACGCGATCTCTACAGCAGGAGGAGGATGCTCAATGCTTACCGGCGAT 549
 QY 541 CGGCAAAACGCGGTTGGCGAATGGAACGCGGCTTAAATTTTCCGCAATCAACTGTCG 600
 DB 550 CGGCAAAACGCGGTTGGCGAATGGAACGCGGCTTAAATTTTCCGCAATCAACTGTCG 609
 QY 601 CTTAAACGTAAGAAACAGAGCAGAAAGCTGTCATTAAGCAGGATTAACCATCGAATC 660
 DB 610 CTTAAACGTAAGAAACAGAGCAGAAAGCTGTCATTAAGCAGGATTAACCATCGAATC 669
 QY 661 AAGTAGAGCGCCGACATGCTCTATTAACCGGTGCGGTAAAGCTCGCATCAATGCTGACG 720
 DB 670 AAGTAGAGCGCCGACATGCTCTATTAACCGGTGCGGTAAAGCTCGCATCAATGCTGACG 729
 QY 721 GAGATATTTCTCTGCAACAGAGCAGGAAATCCGAGCGCGGAGTGGGAGAGATCAC 780
 DB 730 AAGATATTTCTCTGCAACAGAGCAGGAAATCCGAGCGCGGAGTGGGAGAGATCAC 789
 QY 781 GATTCAACCAAGTGAACACCTTGTGAAGTTTGCATTAACCGCAATTTTATTTGTAACA 840
 DB 790 GATTCAACCAAGTGAACACCTTGTGAAGTTTGCATTAACCGCAATTTTATTTGTAACA 849
 QY 841 CGCAGCCAGAGGTTGGCCGACGCGGCGCCACCCGTTATTAGATTGATCAAGACGCG 900
 DB 850 CGCAGCCAGAGGTTGGCCGACGCGGCGCCACCCGTTATTAGATTGATCAAGACGCG 909
 QY 901 TTGAGCGCCGATCAACCGCAAAACAGCGGATGTTGTGACATTAACCATTCAGTCTG 960
 DB 910 TTGAGCGCCGATCAACCGCAAAACAGCGGATGTTGTGACATTAACCATTCAGTCTG 969
 QY 961 TTTATGCGCGGACACGATTAATCTGCGGCAATCTGCGCGGCGACGAGGATCACTG 1020
 DB 970 TTTATGCGCGGACACGATTAATCTGCGGCAATCTGCGCGGCGACGAGGATCACTG 1029
 QY 1021 ACGCTTCCCGGATGAGCGGATTAACGCGCGACAGGTGTAAGTGTGTTGAACGCTG 1080
 DB 1030 ACGCTTCCCGGATGAGCGGATTAACGCGCGACAGGTGTAAGTGTGTTGAACGCTG 1089
 QY 1081 CGTGGGTAAGCAGTAACAGCCAGTGAATCAAGTTTCGCTGTCTTCCAGACTTTACG 1140
 DB 1090 CGTGGGTAAGCAGTAACAGCCAGTGAATCAAGTTTCGCTGTCTTCCAGACTTTACG 1149
 QY 1141 CAGATGCGTAAACAGCGGCTGTCTAATTAATCGCGGCGCGGAGGAGGTAACCTGAC 1200
 DB 1150 CAGATGCGTAAACAGCGGCTGTCTAATTAATCGCGGCGCGGAGGAGGTAACCTGAC 1209
 QY 1201 CTGGCAGAGTGAAGAGCGAAATGCGCAGGAGCATGTTGTTGTTGTTGTTGTTGTTG 1260
 DB 1210 CTGGCAGAGTGAAGAGCGAAATGCGCAGGAGCATGTTGTTGTTGTTGTTGTTGTTG 1269
 QY 1261 ATCGTAATGAAGACGATACCGCTTGCAGTTTGTAA 1299
 DB 1270 ATCGTAATGAAGACGATACCGCTTGCAGTTTGTAA 1308

RESULT 15
 ECOACPHC 1901 bp DNA linear BCT 04-Apr-2002
 LOCUS Escherichia coli periplasmic phosphoanhydride phosphohydrolyase
 DEFINITION (appa) gene, complete cds.
 ACCESSION L03372
 VERSION L03372.1 GI:145180

KEYWORDS acid phosphatase; periplasmic phosphoanhydride phosphohydrolyase.
 SOURCE Escherichia coli
 ORGANISM Escherichia coli
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
 REFERENCE 1 (sites)
 AUTHORS Dassa, J., Marcq, C., and Boquet, P.L.
 TITLE The complete nucleotide sequence of the Escherichia coli gene appA reveals significant homology between pH 2.5 acid phosphatase and glucose-1-phosphatase
 JOURNAL J. Bacteriol. 172 (9), 5497-5500 (1990)
 MEDLINE 90368616
 PUBMED 2168385
 REFERENCE 2 (bases 1 to 1901)
 AUTHORS Oetani, K., Harms, E.H., Stevis, P.E., Kuciel, R., Zhou, M.M. and Van Eten, R.L.
 TITLE Overexpression, site-directed mutagenesis, and mechanism of Escherichia coli acid phosphatase
 JOURNAL J. Biol. Chem. 267 (32), 22830-22836 (1992)
 MEDLINE 93054596
 PUBMED 1429631
 FEATURES
 source
 1..1901
 /organism="Escherichia coli"
 /mol_type="genomic DNA"
 /db_xref="taxon:562"
 188..1486
 /gene="appa"
 188..1486
 /gene="appa"
 /standard_name="acid phosphatase"
 /citation=[1]
 /codon_start=1
 /transl_table=11
 /product="periplasmic phosphoanhydride phosphohydrolyase"
 /protein_id="AA00004.1"
 /db_xref="GI:145181"
 /translation="MKAILIPLSLIPLTPQAFAPQSPBELKESVIVSRGVAP
 TKATQMDVTPDAMPPTWPKGLMLPRGGLIAYIGHYORQIVADGILLAKGCPQ
 GVALIADVDERTKTEAPAGLAPDCAITVHTQDTSPPDLFPLPTGVQQLDNA
 NVYDALISRAAGSIADPTGRQTAFLERLVINFPQSNCLKREKQDESLTQALPS
 ELKESADNYSITGAVALSMLEIFLQQAQWPEQWCRITDSHKQNTLSLHNAQF
 YLQRTPEVARSPPLDLIKTALRHPPOKQAYGVTLPTSLVFLAGHDYTNANLGG
 ALKINWTLPGQDNPPTPGELVPERWRISDQSNQVSLVPTQLQMDKPLSLNT
 PEEVFLTAGCEBNAQKSLAGTQIVNEKRIACSL"
 299..301
 /gene="appa"
 /citation=[1]
 /phenotype="R20A mutant"
 ORIGIN
 Query Match 99.3%; Score 1290; DB 1; Length 1901;
 Best Local Similarity 99.5%; Pred. No. 0;
 Matches 1293; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 variation
 QY 1 ATGAAAGCATCTTAATCCATTTTATCTCTCTGATTCCTGTTAACCCCGGCAATCGCA 60
 DB 188 ATGAAAGCATCTTAATCCATTTTATCTCTCTGATTCCTGTTAACCCCGGCAATCGCA 247
 QY 61 TTGGCTCAGGTAGAGCGGAGCTGAAGCTGGAAGTGTGATGTCAGTCTGATGTT 120
 DB 248 TTGGCTCAGGTAGAGCGGAGCTGAAGCTGGAAGTGTGATGTCAGTCTGATGTT 307
 QY 121 GTGCGTCTCAACCAAGGCGACGCACTGATGACAGATGCAACCCGAGCGATGGCCA 180
 DB 308 GTGCGTCTCAACCAAGGCGACGCACTGATGACAGATGCAACCCGAGCGATGGCCA 367
 QY 181 ACCTGGCCGCTAAACCTGGGTTGGCTGACACCGCGCGGTGTGAGCTAATCGCTATCTC 240
 DB 368 ACCTGGCCGCTAAACCTGGGTTGGCTGACACCGCGCGGTGTGAGCTAATCGCTATCTC 427
 QY 241 GGAATTAACCAAGCCGAGCTGTGTGACGAGATTTGCTGCGAATAAAGGCGTGGCCG 300

428 GGACATTACAAAGCCAGCGCTCTGTAGCCAGCGATTGTGGGAAAAAGGCTGCCG 487
QY 301 CAGTCTGGTCAGAGTCGCGATTATTTGCTGATGTCAAGAGCGTACCCGTTAAAAAGAGGAA 360
DB 488 CAGTCTGGTCAGAGTCGCGATTATTTGCTGATGTCAAGAGCGTACCCGTTAAAAAGAGGAA 547
QY 361 GCGTTGCGCGCGGCGTGGCAGCTGACATTAACGATACCGATACCGAGATACG 420
DB 548 GCGTTGCGCGCGGCGTGGCAGCTGACATTAACGATACCGAGATACCGAGATACG 607
QY 421 TCCAGTCCCGATCCGTTATTTATCTCTTAAAACTGCGGTTTGCCTGCAATGACG 480
DB 608 TCCAGTCCCGATCCGTTATTTATCTCTTAAAACTGCGGTTTGCCTGCAATGACG 667
QY 481 AACGTGACTGACCGGATCTCTGACGAGGAGAGAGGTCATTTGCTGACCTTTACCGGCGAT 540
DB 668 AACGTGACTGACCGGATCTCTGACGAGGAGAGAGGTCATTTGCTGACCTTTACCGGCGAT 727
QY 541 CCGGAAAACGGCGGTTTGGCGAAGCTGAAAGGAGCTTAAATTTTCCGCAATCAAACTTGTGC 600
DB 728 CCGGAAAACGGCGGTTTGGCGAAGCTGAAAGGAGCTTAAATTTTCCGCAATCAAACTTGTGC 787
QY 601 CTTAAAAGTGAAGAAAGAGAGAGAGCTGTTCACTTAAAGGAGCAATGCGAACTC 660
DB 788 CTTAAAAGTGAAGAAAGAGAGAGAGCTGTTCACTTAAAGGAGCAATGCGAACTC 847
QY 661 AAGGTGAGCGCGGCAATGCTCATTTAACGGTGGGTAAAGCTCGCATCAATGCTGACG 720
DB 848 AAGGTGAGCGCGGCAATGCTCATTTAACGGTGGGTAAAGCTCGCATCAATGCTGACG 907
QY 721 GAGATATTTCTCTGCAACAAGCAGAGGAATGCCGAGCGGCGGTGGGAAAGATCAC 780
DB 908 GAGATATTTCTCTGCAACAAGCAGAGGAATGCCGAGCGGCGGTGGGAAAGATCAC 967
QY 781 GATTCAACACAGTGAACAACCTTGTAGTTTGCATTAACGCGCAATTTTATTGCTACAA 840
DB 968 GATTCAACACAGTGAACAACCTTGTAGTTTGCATTAACGCGCAATTTTATTGCTACAA 1027
QY 841 CGCAGCGCAGAGTTGCCGAGCGCGCGCAACCCGTTATTTAGATTGATCAAGACAGCG 900
DB 1028 CGCAGCGCAGAGTTGCCGAGCGCGCGCAACCCGTTATTTAGATTGATCAAGACAGCG 1087
QY 901 TTGAGCGCCCATCCACCGCAAAAAAGAGCGTATGTTGATCAATCCACTTCAGTCTG 960
DB 1088 TTGAGCGCCCATCCACCGCAAAAAAGAGCGTATGTTGATCAATCCACTTCAGTCTG 1147
QY 961 TTTATCGCGCGACAGATATCTATCTGCAAAATCTCGCGCGCACTGAGCTCACTG 1020
DB 1148 TTTATCGCGCGACAGATATCTATCTGCAAAATCTCGCGCGCACTGAGCTCACTG 1207
QY 1021 ACGCTTCCCGGATCAGCGCGATTAACGCGCGCGAGTGTGAACGTGTTGAAGCGCTG 1080
DB 1208 ACGCTTCCCGGATCAGCGCGATTAACGCGCGCGAGTGTGAACGTGTTGAAGCGCTG 1267
QY 1081 CCGCGGCTTAAGGATTAACAGCAGTGAATTCAGTTTCCGCTGCTTCAGACTTTACG 1140
DB 1268 CCGCGGCTTAAGGATTAACAGCAGTGAATTCAGTTTCCGCTGCTTCAGACTTTACG 1327
QY 1141 CAGATCGTGAATTAACGCGCGCTGCTATTAATACGCGCGCGAGAGGTGAACCTGACC 1200
DB 1328 CAGATCGTGAATTAACGCGCGCTGCTATTAATACGCGCGCGAGAGGTGAACCTGACC 1387
QY 1201 CTGGCAGAGTGTGAAGAGGAAATGGCAGGGCATGTTCGTTGGCAGGTTTACGCA 1260
DB 1388 CTGGCAGAGTGTGAAGAGGAAATGGCAGGGCATGTTCGTTGGCAGGTTTACGCA 1447
QY 1261 ATCGTGAATGAAGACGATACCGCTGCGCTGCAATTTGTAA 1299
DB 1448 ATCGTGAATGAAGACGATACCGCTGCGCTGCAATTTGTAA 1486

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 20:12:02 / Search time 885.787 Seconds
(without alignments)
8681.254 Million cell updates/sec

Title: US-09-926-375B-7_COPY_12653_13951

Perfect score: 1299

Sequence: 1 atgaagcattcttaatccc.....taccgcgttcgactttgtta 1299

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 4390206 segs, 2959870667 residues 8780412

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: geneseqn19808:*
2: geneseqn19908:*
3: geneseqn20008:*
4: geneseqn20018:*
5: geneseqn20028:*
6: geneseqn20038:*
7: geneseqn20048:*
8: geneseqn20058:*
9: geneseqn20068:*
10: geneseqn20078:*
11: geneseqn20088:*
12: geneseqn20098:*
13: geneseqn20108:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1299	100.0	3470	3	AAC68298	AAC68298 R15/APPA
2	1299	100.0	4060	3	AAC68296	AAC68296 R15/APPA
3	1299	100.0	6116	3	AAC68297	AAC68297 R15/APPA
4	1299	100.0	6708	3	AAC68295	AAC68295 R15/APPA
5	1299	100.0	17732	3	AAC68300	AAC68300 Lama2/APP
6	1299	100.0	20623	3	AAC68294	AAC68294 Lama2/APP
7	1294.2	99.6	1299	3	ABK12514	ABK12514 DNA encod
8	1294.2	99.6	1299	3	ABK12514	ABK12514 DNA encod
9	1294.2	99.6	1299	3	ABK12514	ABK12514 DNA encod
10	1293.2	99.6	1299	3	ADL16137	ADL16137 Escherich
11	1293.2	99.6	1299	3	ADL16137	ADL16137 Escherich
12	1293.2	99.6	1299	3	ADL16137	ADL16137 Escherich
13	1293.2	99.6	1299	3	ADL16137	ADL16137 Escherich
14	1290	99.3	1901	12	ADOS0295	ADOS0295 Escherich
15	1288.4	99.2	1296	12	ADOS0296	ADOS0296 Escherich
16	1288	99.2	1296	12	ADL16139	ADL16139 Shigella
17	1287.8	99.1	1315	10	ADL02197	ADL02197 DNA encod
18	1286	98.7	5421	3	AAC68299	AAC68299 SV40/APPA
19	1282.2	98.7	1323	5	AAC88885	AAC88885 Escherich
20	1282.2	98.7	1323	6	AAD36473	AAD36473 Escherich

21	1282.2	98.7	1323	6	AAD25460	AAD25460 Escherich
22	1282.2	98.7	1323	8	ADA19445	ADA19445 E. coli B
23	1282.2	98.7	1323	12	ADOS0291	ADOS0291 Escherich
24	1282.2	98.7	1323	12	ADOS0301	ADOS0301 Kanaroc
25	1282.2	98.7	1323	12	ADOS0303	ADOS0303 Escherich
26	1268.8	97.7	1296	12	ADL16138	ADL16138 Shigella
27	1267.6	97.6	1901	8	ADA19452	ADA19452 E. coli K
28	1263.8	97.3	1486	4	AAD06831	AAD06831 E. coli a
29	1259.2	96.9	1108	10	ADC87742	ADC87742 DNA encod
30	1259.2	96.9	1308	12	ADOS0299	ADOS0299 Escherich
31	1257.4	96.8	1489	3	AAA28216	AAA28216 E. coli a
32	1257.4	96.8	1489	8	ACC57672	ACC57672 Escherich
33	1252.6	96.4	1486	4	AAD06832	AAD06832 E. coli a
34	1252.6	96.4	1486	4	ACC57673	ACC57673 Escherich
35	1170.2	90.1	1272	2	AAK26540	AAK26540 DNA encod
36	861.6	66.3	868	12	ADL16143	ADL16143 Proteus v
37	819.8	63.1	846	4	AAK1940	AAK1940 Genomic s
38	750.8	57.8	754	12	ADL16144	ADL16144 Zoo compo
39	749	57.7	757	12	ADL16142	ADL16142 Enterobac
40	744.2	57.3	757	12	ADL16141	ADL16141 Enterobac
41	737.2	56.8	1256	9	AAD57147	AAD57147 Maize-opt
42	735.2	56.6	1313	9	AAD57148	AAD57148 PNOV4054
43	734.2	56.5	1331	9	AAD57149	AAD57149 PNOV4058
44	722.4	55.6	724	12	ADL16140	ADL16140 Pasteurel
45	694.8	53.5	1737	5	AAS89885	AAS89885 DNA encod

ALIGNMENTS

RESULT 1	
AAC68298	AAC68298 standard; DNA; 3470 BP.
XX	XX
AC	AAC68298;
XX	XX
DT	15-SRP-2003 (revised)
DT	20-FEB-2001 (first entry)
XX	XX
DE	R15/APPA plasmid coding sequence.
XX	XX
KW	Transgenic animal; salivary protein; phytase; phosphorus; animal growth;
KW	environmental pollution; pig; ds.
OS	Rattus sp.
OS	Escherichia coli.
OS	Chimeric.
XX	XX
PN	WO200064247-A1.
XX	XX
XX	02-NOV-2000.
XX	XX
PD	20-APR-2000; 2000MO-CA000430.
XX	XX
PR	23-APR-1999; 99US-0130508P.
XX	XX
PA	(UYGU-) UNIV GUELPH.
XX	XX
PI	Forsberg CW, Golovan S, Phillips JP;
PI	WPI; 2000-687245/67.
DR	P-PSDB; AAB36261.
XX	XX
PT	Transgenic non-human animal for gastrointestinal tract specific
PT	expression of a protein, preferably phytase, comprises a nucleic acid
PT	sequence including a heterologous transgene construct encoding the
PT	protein.
XX	XX
PS	Claim 14; Fig 21; 152pp; English.
XX	XX
CC	The present invention provides transgenic animals which produce desired
CC	proteins, in this case pigs which express phytase in the salivary
CC	gland. Low phytase production levels result in phytase in the diet being

CC excreted and causing phosphorus contamination in water, as well as
CC reducing the growth of animals. The invention provides a number of
CC transgenes containing the *E. coli* APHA phylatae coding sequence. (Updated
CC on 15-SEP-2003 to standardise OS field)
XX
SQ Sequence 3470 BP; 1065 A; 721 C; 735 G; 949 T; 0 U; 0 Other;

Sequence 3470 BP; 1065 A; 721 C; 735 G; 949 T; 0 U; 0 Other;

Query Match	100.0%	Score 1299;	DB 3,	Length 3470;
Best Local Similarity	100.0%;	Pred. NC. 0;		
Matches 1299; Conservative	0;	Mismatches	0;	Gaps 0

QY	1	TTGAAAGCATCTTAAATCCCAATTTTAACTCTCTGATTTCCGTTAAACCCGCAATCTGA	60
Db	1811	ATGAAAGCATCTTAAATCCCAATTTTAACTCTCTGATTTCCGTTAAACCCGCAATCTGA	1870
QY	61	TTCCGCTCAGAGTGAAGCCGAGCTGAAGCTGAAAGTGTGTGATTTGTCAGTGTCATGAT	120
Db	1871	TTCCGCTCAGAGTGAAGCCGAGCTGAAGCTGAAAGTGTGTGATTTGTCAGTGTCATGAT	193
QY	121	GTGCGTGCTCCAACCAAGGCCACGCAACTGATGCAAGATGTATACCCCAAGCCATGGCCA	180
Db	1931	GTGCGTGCTCCAACCAAGGCCACGCAACTGATGCAAGATGTATACCCCAAGCCATGGCCA	1990
QY	181	ACCTGGCCGGTAAACCTGGATTGGCTGACACCGCCGCGGTGGAGCTAATCGCTATCTC	240
Db	1991	ACCTGGCCGGTAAACCTGGATTGGCTGACACCGCCGCGGTGGAGCTAATCGCTATCTC	205
QY	241	GGACATTAACCAAGCCAGGTCGTGTAAGCCGAAGGATTTGCGGAAAAAGGCTGGCCG	300
Db	2051	GGACATTAACCAAGCCAGGTCGTGTAAGCCGAAGGATTTGCGGAAAAAGGCTGGCCG	211
QY	301	CAGTCTGGTCAAGTGGCCATTATTTGCTGATGTCTGACGAGCGTAAACCCGTAAACAGGCCGA	360
Db	2111	CAGTCTGGTCAAGTGGCCATTATTTGCTGATGTCTGACGAGCGTAAACCCGTAAACAGGCCGA	217
QY	361	GCCTTGCGCGCCGGGCTGGCACCCTGACTGTGATAATACCGTATACCAGGCAAGATACG	420
Db	2171	GCCTTGCGCGCCGGGCTGGCACCCTGACTGTGATAATACCGTATACCAGGCAAGATACG	223
QY	421	TCGAGTCCCGATCGCTTAATTATTCCTCTAAAACTGCGCTTTGCCACTGATTAACGCG	480
Db	2231	TCGAGTCCCGATCGCTTAATTATTCCTCTAAAACTGCGCTTTGCCACTGATTAACGCG	229
QY	481	AAAGTGAATGAAGGCATCTCTGACGAGGCGAGAGGGTCAATTGTGATCTTACCGGGCAT	540
Db	2291	AAAGTGAATGAAGGCATCTCTGACGAGGCGAGAGGGTCAATTGTGATCTTACCGGGCAT	235
QY	541	CGGCAAAAGGCGCTTTTCGCGAACTGGAACGGGTGCTTAATTTTCGCGAATCAAACTTGTGC	600
Db	2351	CGGCAAAAGGCGCTTTTCGCGAACTGGAACGGGTGCTTAATTTTCGCGAATCAAACTTGTGC	241
QY	601	CTTAAACGTGAGAAACAGACGAAAGCTTTTCAATTAACGCAAGCATTAACATCGGAATCTC	660
Db	2411	CTTAAACGTGAGAAACAGACGAAAGCTTTTCAATTAACGCAAGCATTAACATCGGAATCTC	247
QY	661	AAGGAGAGGCGCGACATGTCTCATTAACGGGAGCGGTAAAGCTCGCATCAATGTGACG	720
Db	2471	AAGGAGAGGCGCGACATGTCTCATTAACGGGAGCGGTAAAGCTCGCATCAATGTGACG	253
QY	721	GAGATATTTCTCTGCAACAGCAACAGGAATGCCGAAGCCGGGAGTGGGAAAGATCACT	780
Db	2531	GAGATATTTCTCTGCAACAGCAACAGGAATGCCGAAGCCGGGAGTGGGAAAGATCACT	259
QY	781	GATTCAACCAAGTGGAAACCTTGTCTAAATTGGCAATAAAGCCCAATTTTATTTGCTACAA	840
Db	2591	GATTCAACCAAGTGGAAACCTTGTCTAAATTGGCAATAAAGCCCAATTTTATTTGCTACAA	265
QY	841	CGCAGCGCAGAGTGTCCCGCAGCCGCGCACCCGCTTATTAAGATTGATTAACAAGCAGCG	900
Db	2651	CGCAGCGCAGAGTGTCCCGCAGCCGCGCACCCGCTTATTAAGATTGATTAACAAGCAGCG	271
QY	901	TTGACGCCCATTCACCGCAAAAAACAGGCGTATGTGTGACATTAACCACTTTCAGTGTG	960

CC on 15-SEP-2003 to standardise OS field)

XX Sequence 4060 BP; 1257 A; 814 C; 843 G; 1146 T; 0 U; 0 Other;

SQL Query Match 100.0%; Score 1299; DB 3; Length 4060;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGAAAGCATCTTAAATCCCATTTTATCTCTTCTGATTCCTGTAACCCCGCAATCTGCA 60
 DB 1811 ATGAAAGCATCTTAAATCCCATTTTATCTCTTCTGATTCCTGTAACCCCGCAATCTGCA 1870

QY 61 TTGCTCAAGATGAGCCGAGCTGAAAGCTGGAAGTGTGTGATTTGTCACTGTCTGTGT 120
 DB 1871 TTGCTCAAGATGAGCCGAGCTGAAAGCTGGAAGTGTGTGATTTGTCACTGTCTGTGT 1930

QY 121 GTGCGGTCTCAACCAAGGCGACGCAATGATGAGATGTCACCCGAGAGCATGAGCCCA 180
 DB 1931 GTGCGGTCTCAACCAAGGCGACGCAATGATGAGATGTCACCCGAGAGCATGAGCCCA 1990

QY 181 ACCTGACCGGTAAACCTGAGTGTGCTGACACCGCGGTGTGATGCTAATCGCTATCTC 240
 DB 1991 ACCTGACCGGTAAACCTGAGTGTGCTGACACCGCGGTGTGATGCTAATCGCTATCTC 2050

QY 241 GGAATTTACCAACGCCAGCTCTGTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 300
 DB 2051 GGAATTTACCAACGCCAGCTCTGTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 2110

QY 301 CAGTCTGTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 360
 DB 2111 CAGTCTGTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 2170

QY 361 GCCTTGGCCGCGCGGCTGACCTGACCTGACCTGACCTGACCTGACCTGACCTGACCTGAC 420
 DB 2171 GCCTTGGCCGCGCGGCTGACCTGACCTGACCTGACCTGACCTGACCTGACCTGACCTGAC 2230

QY 421 TCCAGTCCCGATCCGTTATTTAATCTCTAATAAATCTGCGTTTCCCACTGATTAACGCG 480
 DB 2231 TCCAGTCCCGATCCGTTATTTAATCTCTAATAAATCTGCGTTTCCCACTGATTAACGCG 2290

QY 481 AAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 540
 DB 2291 AAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2350

QY 541 CGGCAAAACGCGCTTTCCGAACTGGAACGCGGTCTTAATTTTCCGCAATCAAACTTGTGC 600
 DB 2351 CGGCAAAACGCGCTTTCCGAACTGGAACGCGGTCTTAATTTTCCGCAATCAAACTTGTGC 2410

QY 601 CTTAAACGTGAGAAACGAGACGAAAGCTGTTCAATTAACGAGCATTAACATCGGAATCTC 660
 DB 2411 CTTAAACGTGAGAAACGAGACGAAAGCTGTTCAATTAACGAGCATTAACATCGGAATCTC 2470

QY 661 AAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 720
 DB 2471 AAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2530

QY 721 GAGATATTTTCTCTGCAACAAAGCAGAGGAATGCGGAGCGGGGTGGGAGGATCAAC 780
 DB 2531 GAGATATTTTCTCTGCAACAAAGCAGAGGAATGCGGAGCGGGGTGGGAGGATCAAC 2590

QY 781 GATTGACACCAAGTGAACACCTTGTCTAAGTTTGATTAACGCGCAATTTTATTTGCTCA 840
 DB 2591 GATTGACACCAAGTGAACACCTTGTCTAAGTTTGATTAACGCGCAATTTTATTTGCTCA 2650

QY 841 CGGACGCGCAGAGTGTGCGCGGCGGCGCAACCCGTTAATTAAGTTTGAATTAAGACGCG 900
 DB 2651 CGGACGCGCAGAGTGTGCGCGGCGGCGCAACCCGTTAATTAAGTTTGAATTAAGACGCG 2710

QY 901 TTGACGCGCCATCCACCGCAAAACAGGCGTATGATGATTAACCATTTCACTTCACTGTG 960
 DB 2711 TTGACGCGCCATCCACCGCAAAACAGGCGTATGATGATTAACCATTTCACTTCACTGTG 2770

QY 961 TTTATCGCGGACACGATTAATCTGCAAAATCTCGGCGGCGCATGAGACTCAACTGG 1020

DB 2771 TTTATCGCGGACACGATTAATCTGCAAAATCTCGGCGGCGCATGAGACTCAACTGG 2830

QY 1021 ACGCTTCCCGGTGACGCGGATTAACACGCGCCAGAGTGTGAACTGTGTTTGAACGCTGG 1080
 DB 2831 ACGCTTCCCGGTGACGCGGATTAACACGCGCCAGAGTGTGAACTGTGTTTGAACGCTGG 2890

QY 1081 CGTCCGCTAAGGATTAACAGCCAGTGGATTAAGTTCAGTTTGTGCTGCTTCCAGACTTACAG 1140
 DB 2891 CGTCCGCTAAGGATTAACAGCCAGTGGATTAAGTTCAGTTTGTGCTGCTTCCAGACTTACAG 2950

QY 1141 CAGATCGTGTATTAACCGCGCTGTCTAATTAATACCGCCCGGAGAGTGAATGATGAC 1200
 DB 2951 CAGATCGTGTATTAACCGCGCTGTCTAATTAATACCGCCCGGAGAGTGAATGATGAC 3010

QY 1201 CTGGCAGATGTGAAGAGGAAATGCGACAGGAGATGTTGTTGCGAGGTTTACGCA 1260
 DB 3011 CTGGCAGATGTGAAGAGGAAATGCGACAGGAGATGTTGTTGCGAGGTTTACGCA 3070

QY 1261 ATGTGATGAAGCAGCATACCCGCTTGCACTTGTAA 1299
 DB 3071 ATGTGATGAAGCAGCATACCCGCTTGCACTTGTAA 3109

RESULT 3
 AAC68297
 ID AAC68297 standard; DNA: 6116 BP.
 XX AAC68297;
 AC 15-SBP-2003 (revised)
 DT 20-FEB-2001 (first entry)
 XX
 DE R15/APPA plasmid coding sequence.
 XX
 KW Transgenic animal; salivary protein; phytase; phosphorus; animal growth;
 XX environmental pollution; pig; de.
 XX
 OS Rattus sp.
 OS Escherichia coli.
 OS Chimeric.
 XX
 PN M0200064247-A1.
 PD 02-NOV-2000.
 XX
 PF 20-APR-2000; 2000MO-CA000430.
 XX
 PR 23-APR-1999; 99US-0130508P.
 XX
 PA (UYGU-) UNIV GUELPH.
 XX
 PI Forsberg CW, Golovan S, Phillips JP;
 DR WPI, 2000-687245/67.
 DR P-PSDB; AAB36260.
 XX
 PT Transgenic non-human animal for gastrointestinal tract specific
 PT expression of a protein, preferably phytase, comprises a nucleic acid
 PT sequence including a heterologous transgene construct encoding the
 PT protein.
 XX
 PS Claim 56; Fig 20; 152pp; English.
 XX
 CC The present invention provides transgenic animals which produce desired
 CC proteins, in this case pigs which expresses phytase in the salivary
 CC gland. Low phytase production levels result in phytate in the diet being
 CC excreted and causing phosphorus contamination in water, as well as
 CC reducing the growth of animals. The invention provides a number of
 CC transgenes containing the E. coli APPA phytase coding sequence. (Updated
 CC on 15-SEP-2003 to standardise OS field)

XX Sequence 6116 BP; 1724 A; 1386 C; 1407 G; 1599 T; 0 U; 0 Other;

Query Match 100.0%; Score 1299; DB 3; Length 6116;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGAAGCATTTATCCATTTTATCTCTGATTCGGTAAACCCGCAATCTGCA 60
 1811 ATGAAGCATTTATCCATTTTATCTCTGATTCGGTAAACCCGCAATCTGCA 1870
 61 TTCGTCAGAGTAGACCCGAGCTGAAGCTGAAAGTGTGATTTGTCAGTGTATGCT 120
 1871 TTCGTCAGAGTAGACCCGAGCTGAAGCTGAAAGTGTGATTTGTCAGTGTATGCT 1930
 121 GTGCGTCTCCACCAAGCCGCACTGATGCAAGATGTACATCCCGAGAGCTGGCA 180
 1931 GTGCGTCTCCACCAAGCCGCACTGATGCAAGATGTACATCCCGAGAGCTGGCA 1990
 181 ACTGAGCGGTTAAACTGGGTTGGGTGACACCGCGGCGGTGATGAGCTAATGCGCTATCTC 240
 1991 ACTGAGCGGTTAAACTGGGTTGGGTGACACCGCGGCGGTGATGAGCTAATGCGCTATCTC 2050
 241 GGACATTAACCAAGCCGAGCTGTGATGCGCAAGATTGCTGGCGAAGGAGCTGCCG 300
 2051 GGACATTAACCAAGCCGAGCTGTGATGCGCAAGATTGCTGGCGAAGGAGCTGCCG 2110
 301 CAGTCTGTGATGAGCTGTGATTTATGCTGATGTGCAAGACCTGATACCGGTAAACAGGCCGA 360
 2111 CAGTCTGTGATGAGCTGTGATTTATGCTGATGTGCAAGACCTGATACCGGTAAACAGGCCGA 2170
 361 GCGTTCGCGCGCGGGGTGCACTGATGCAATTAACGATACCTGAGAGCTGAGTACG 420
 2171 GCGTTCGCGCGGGGTGCACTGATGCAATTAACGATACCTGAGAGCTGAGTACG 2230
 421 TCCAGTCCCGATCCGTTATTTAATCTCTGATTAATCTGCTGATTAATCTGCTGATTAATCTGCTG 480
 2231 TCCAGTCCCGATCCGTTATTTAATCTCTGATTAATCTGCTGATTAATCTGCTGATTAATCTGCTG 2290
 481 AACGTACTGACGCGCATCTCTGACAGAGGAGGATGCAATTTGCTGATTAATCTGCTGATTAATCTGCTG 540
 2291 AACGTACTGACGCGCATCTCTGACAGAGGAGGATGCAATTTGCTGATTAATCTGCTGATTAATCTGCTG 2350
 541 CGGCAAAAGCGGCTTTGGCGAATGGAACGAGGATCTTAATTTTCCGCAATCAAACTTGTGC 600
 2351 CGGCAAAAGCGGCTTTGGCGAATGGAACGAGGATCTTAATTTTCCGCAATCAAACTTGTGC 2410
 601 CTTAAACGTGAGAAACAGAGCAAGAACTGTCTATTACGCAAGCATTAACCATCGAACTC 660
 2411 CTTAAACGTGAGAAACAGAGCAAGAACTGTCTATTACGCAAGCATTAACCATCGAACTC 2470
 661 AAGGTAGAGCGCGCAATGTCTCTATTAAACCGGTGCGGTAAAGCTCGCATCAATGCTGAGG 720
 2471 AAGGTAGAGCGCGCAATGTCTCTATTAAACCGGTGCGGTAAAGCTCGCATCAATGCTGAGG 2530
 721 GAGATATTTCTCTGCAACAGAGGAAATCCGAGCGCGGATGCGGAGGAGGATCAAC 780
 2531 GAGATATTTCTCTGCAACAGAGGAAATCCGAGCGCGGATGCGGAGGAGGATCAAC 2590
 781 GATTCAACACGATGAAACCTTGTAGTTGCAATTAACGCGCAATTTTATTTGCTACAA 840
 2591 GATTCAACACGATGAAACCTTGTAGTTGCAATTAACGCGCAATTTTATTTGCTACAA 2650
 841 CGACGCCGAGAGGTTCGCGGAGCGCGCAACCCGCTTATTAGATTTGATCAAGACAGG 900
 2651 CGACGCCGAGAGGTTCGCGGAGCGCGCAACCCGCTTATTAGATTTGATCAAGACAGG 2710
 901 TTGACGCCCATCATCGCAAAAGAGGAGGATGATGATGATTAACCATTAAGTGTG 960
 2711 TTGACGCCCATCATCGCAAAAGAGGAGGATGATGATGATTAACCATTAAGTGTG 2770
 961 TTTATGCGCGGACAGATTAATTAATCTGCAAAATCTGCGCGGCGCACTGAGCTCAACTG 1020
 2771 TTTATGCGCGGACAGATTAATTAATCTGCAAAATCTGCGCGGCGCACTGAGCTCAACTG 2830

1021 ACGCTTCCCGGTCAAGCCGATTAACACGCGCAGAGTGTGATGCTGTTGAAAGCTG 1080
 2831 ACGCTTCCCGGTCAAGCCGATTAACACGCGCAGAGTGTGATGCTGTTGAAAGCTG 2890
 1081 CGTGGCTTAACCGATTAACAGCCAGTGTGATGATTTGCTGCTGCTTCCAGACTTTACG 1140
 2891 CGTGGCTTAACCGATTAACAGCCAGTGTGATGATTTGCTGCTGCTTCCAGACTTTACG 2950
 1141 CAGATGCGTGTAAACCGCGCTGTCAATTAATCGCGCGCGGAGAGGTGAACTGACC 1200
 2951 CAGATGCGTGTAAACCGCGCTGTCAATTAATCGCGCGCGGAGAGGTGAACTGACC 3010
 1201 CTGGCAGATGTGAAGACGAATGCGAGGAGCATGTTGCTGTCAGGTTTAAACGCA 1260
 3011 CTGGCAGATGTGAAGACGAATGCGAGGAGCATGTTGCTGTCAGGTTTAAACGCA 3070
 1261 ATCGTGAATGAAGACGATTAACCGCTTGCAGTTTGTAA 1299
 3071 ATCGTGAATGAAGACGATTAACCGCTTGCAGTTTGTAA 3109

RESULT 4
 AAC68295
 ID AAC68295 standard; DNA; 6708 BP.
 XX
 AC AAC68295;
 XX
 DT 15-SEP-2003 (revised)
 DT 20-FEB-2001 (first entry)
 XX
 DE R15/APPA plasmid coding sequence.
 XX
 KM Transgenic animal; salivary protein; phytase; phosphorus; animal growth;
 KM environmental pollution; pig; ds.
 XX
 OS Rattus sp.
 OS Escherichia coli.
 OS Chimeric.
 PN WO20064247-A1.
 XX
 XX
 PD 02-NOV-2000.
 XX
 PF 20-APR-2000; 2006WO-CA000430.
 XX
 PR 23-APR-1999; 99US-0130508P.
 XX
 PA (UYGU-) UNIV GUELPH.
 XX
 PI Forsberg CW, Golovan S, Phillips JP;
 XX
 DR WPI; 2000-687245/67.
 XX
 PS P-PSDB; AAB36258.
 XX
 PT Transgenic non-human animal for gastrointestinal tract specific
 PT expression of a protein, preferably phytase, comprises a nucleic acid
 PT sequence including a heterologous transgene construct encoding the
 PT protein.
 XX
 PS Claim 56; Fig 18; 1522P; English.
 XX
 CC The present invention provides transgenic animals which produce desired
 CC proteins, in this case pigs which expresses phytase in the salivary
 CC gland. Low phytase production levels result in phytate in the diet being
 CC excreted and causing phosphorus contamination in water, as well as
 CC reducing the growth of animals. The invention provides a number of
 CC transgenes containing the E. coli APPA phytase coding sequence. (Updated
 CC on 15-SEP-2003 to standardise OS field)
 XX
 SQ Sequence 6708 BP; 1916 A; 1479 C; 1515 G; 1798 T; 0 U; 0 Other;

Query Match 100.0%; Score 1299; DB 3; Length 6708;
 Best Local Similarity 100.0%; Pred. No. 0;

Matches 1299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAAAGCCATCTTAATCCCATTTTATCTTTCTGATTCGGTAAACCCCGCAATCTGCA 60
DB 1811 ATGAAAGCCATCTTAATCCCATTTTATCTTTCTGATTCGGTAAACCCCGCAATCTGCA 1870

QY 61 TTGCGCTCAAGTGAAGCCGAGCTGAAAGCTGGAAGAGTGTGATTTTCTGCTGCTCATGTGT 120
DB 1871 TTGCGCTCAAGTGAAGCCGAGCTGAAAGAGTGTGATTTTCTGCTGCTCATGTGT 1930

QY 121 GTCGCGTCTCAACCAAGGCGACGCAACTGATSCAGATGTCAACCCAGACGATGGCA 180
DB 1931 GTCGCGTCTCAACCAAGGCGACGCAACTGATSCAGATGTCAACCCAGACGATGGCA 1990

QY 181 ACCTGGCCGGTAAACCTGGGTTGGCTGACACCGCGCGGTGGTGAAGCTAATCGCTAATCTC 240
DB 1991 ACCTGGCCGGTAAACCTGGGTTGGCTGACACCGCGCGGTGGTGAAGCTAATCGCTAATCTC 2050

QY 241 GGAACATTACCAAGCCAGCGTCTGTGTAAGCCGACGATTTGTCGGCGAATAAAGGCGTCCCG 300
DB 2051 GGAACATTACCAAGCCAGCGTCTGTGTAAGCCGACGATTTGTCGGCGAATAAAGGCGTCCCG 2110

QY 301 CAGCTGTCAGGTCGGGATTATTGCTGATGTGACGAGCGGTACCGGTAAACAGCGCA 360
DB 2111 CAGCTGTCAGGTCGGGATTATTGCTGATGTGACGAGCGGTACCGGTAAACAGCGCA 2170

QY 361 GCGCTTCGCGCGCGGGCTGCGACCTGACTGTGCAATAACCGTACATACCCAGGCAATACG 420
DB 2171 GCGCTTCGCGCGCGGGCTGCGACCTGACTGTGCAATAACCGTACATACCCAGGCAATACG 2230

QY 421 TCCAGTTCGCGATCCGTTATTTAATCTCTTAAAACTGGCTTGGCCAACTGATTAACGCG 480
DB 2231 TCCAGTTCGCGATCCGTTATTTAATCTCTTAAAACTGGCTTGGCCAACTGATTAACGCG 2290

QY 481 AAGGTACGTACGCGGATCTCAGAGAGGAGAGAGGCTCAATTGCTGACTTTACCGGCGAT 540
DB 2291 AAGGTACGTACGCGGATCTCAGAGAGGAGAGAGGCTCAATTGCTGACTTTACCGGCGAT 2350

QY 541 CGGCAAAACGCGCTTTTCGCAACTGGAACGGGTGCTTAATTTTCGCAATCAAACTTGTGC 600
DB 2351 CGGCAAAACGCGCTTTTCGCAACTGGAACGGGTGCTTAATTTTCGCAATCAAACTTGTGC 2410

QY 601 CTTAAACGTGAAACAGACGAAAGCTGTTCAATTAACGAGGATTAACCATGGAATCTC 660
DB 2411 CTTAAACGTGAAACAGACGAAAGCTGTTCAATTAACGAGGATTAACCATGGAATCTC 2470

QY 661 AAGGTGAGCGCGCAATGTCTCATTAACGGGTGCGTAAAGCCTCGCATCAATGCTGACG 720
DB 2471 AAGGTGAGCGCGCAATGTCTCATTAACGGGTGCGTAAAGCCTCGCATCAATGCTGACG 2530

QY 721 GAGATATTTCTCTGCAACAGCAACAGGAATGCCGAGTCCGGGTGGGAGAGATCAC 780
DB 2531 GAGATATTTCTCTGCAACAGCAACAGGAATGCCGAGTCCGGGTGGGAGAGATCAC 2590

QY 781 GATTACACCAAGTGAACACCTTGTCTAAGTTGATTAACGGCAATTTTATTTGCTACAA 840
DB 2591 GATTACACCAAGTGAACACCTTGTCTAAGTTGATTAACGGCAATTTTATTTGCTACAA 2650

QY 841 CGACGCGCAGAGGTGGCCGAGCGCGGCAACCCGTTATTAAGTTGATCAAGACGCG 900
DB 2651 CGACGCGCAGAGGTGGCCGAGCGCGGCAACCCGTTATTAAGTTGATCAAGACGCG 2710

QY 901 TTGACGCGCCATCCACCGCAAAAACAGGCGTATGTGATTAACCATCTTCAGTCTG 960
DB 2711 TTGACGCGCCATCCACCGCAAAAACAGGCGTATGTGATTAACCATCTTCAGTCTG 2770

QY 961 TTTATGCGCGGACACGATTAATCTGCAAAATCTCGCGCGCGCATGAGACTCAACTGG 1020
DB 2771 TTTATGCGCGGACACGATTAATCTGCAAAATCTCGCGCGCGCATGAGACTCAACTGG 2830

QY 1021 ACGCTTCGCGGTACGCGGATTAACGCGCGCAGGTGTGTAAGTGTGTTGAACGCTGG 1080
DB 2831 ACGCTTCGCGGTACGCGGATTAACGCGCGCAGGTGTGTAAGTGTGTTGAACGCTGG 2890

QY 1081 CGTCGGCTAAGGATTAACAGCCAGTGGATTGAGTTTGGCTGATCTTCAGACTTTACAG 1140
DB 2891 CGTCGGCTAAGGATTAACAGCCAGTGGATTGAGTTTGGCTGATCTTCAGACTTTACAG 2950

QY 1141 CAGATGCGTGAATAAACCGCGCTGTCTAATTAATAGCCCGCGAGAGGTGAACGTAGC 1200
DB 2951 CAGATGCGTGAATAAACCGCGCTGTCTAATTAATAGCCCGCGAGAGGTGAACGTAGC 3010

QY 1201 CTGGCAGATGTGAAGAGCGAAATGCGCAGGCGATGTGTTGTTGGCAGTTTACGCA 1260
DB 3011 CTGGCAGATGTGAAGAGCGAAATGCGCAGGCGATGTGTTGTTGGCAGTTTACGCA 3070

QY 1261 ATCGTGAATGAAGACGCAATACCCGCTGACGTTGTAA 1299
DB 3071 ATCGTGAATGAAGACGCAATACCCGCTGACGTTGTAA 3109

RESULT 5
AAC68300
ID AAC68300 standard; DNA; 17732 BP.
AC AAC68300;
XX
XX
DT 15-SBP-2003 (revised)
DT 20-FEB-2001 (first entry)
XX
XX
DE Lama2/APPA plasmid coding sequence.
XX
XX Transgenic animal; salivary protein; phytase; phosphorus; animal growth;
KM environmental pollution; pig; de.
XX
XX Mus musculus.
OS Escherichia coli.
OS Chimeric.
XX
PN WO200064247-A1.
XX
PD 02-NOV-2000.
XX
PF 20-APR-2000; 2000WO-CA000430.
XX
PR 23-APR-1999; 99US-0130508P.
XX
XX (UYGU-) UNIV GUELPH.
XX
XX Forsberg CW, Golovan S, Phillips JP;
PI WPI; 2000-687245/67.
DR P-PSDB; AAB36263.
XX
XX Transgenic non-human animal for gastrointestinal tract specific
PT expression of a protein, preferably phytase, comprises a nucleic acid
PT sequence including a heterologous transgene construct encoding the
PT protein.
XX
XX Claim 14; Fig 23; 152pp; English.
XX
XX The present invention provides transgenic animals which produce desired
CC proteins, in this case pigs which expresses phytase in the salivary
CC gland. Low phytase production levels result in phytate in the diet being
CC excreted and causing phosphorus contamination in water, as well as
CC reducing the growth of animals. The invention provides a number of
CC transgenes containing the E. coli APPA phytase coding sequence. (Updated
CC on 15-SBP-2003 to standardise OS field)
XX
SQ Sequence 17732 BP; 4719 A; 4125 C; 4168 G; 4719 T; 0 U; 1 Other;
Query Match 100.0%; Score 1299; DB 3; Length 17732;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAAAGCCATCTTAATCCCATTTTATCTTTCTGATTCGGTAAACCCCGCAATCTGCA 60

```

Db 12653 ATGAAAGCATCTTAATCCATTTTATCTCTTCGATTCCTGTTAACCCCGCAATCTGCA 12712
Qy 61 TTGCTGCAAGTAGAGCCGAGCTGAAAGCTGAAAGTGTGTGATTTGTGATGTCATGAT 120
Db 12713 TTGCTGCAAGTAGAGCCGAGCTGAAAGCTGAAAGTGTGTGATTTGTGATGTCATGAT 12772
Qy 121 GTGCGTCTCCAAACCAAGCCCACTGATGCAAGATGTCAACCCCAAGCCGATGGCA 180
Db 12773 GTGCGTCTCCAAACCAAGCCCACTGATGCAAGATGTCAACCCCAAGCCGATGGCA 12832
Qy 181 ACCCGCCGCTAAACTGCGTGGCTGACACCGCGGTGTGAGCTAATCGCTATCTC 240
Db 12833 ACCCGCCGCTAAACTGCGTGGCTGACACCGCGGTGTGAGCTAATCGCTATCTC 12892
Qy 241 GGACATTTACCAAGCCGCTGTGTAGCCGACGGAATTCGCGCAAAAAGGCTGCCG 300
Db 12893 GGACATTTACCAAGCCGCTGTGTAGCCGACGGAATTCGCGCAAAAAGGCTGCCG 12952
Qy 301 CAGTCTGCTCAGGTCCGCTATTTATGCTGATGTCAGACGACGTAACCCGTAACCAAGCCGA 360
Db 12953 CAGTCTGCTCAGGTCCGCTATTTATGCTGATGTCAGACGACGTAACCCGTAACCAAGCCGA 13012
Qy 361 GCCTTCCGCGCGGCTGACCTGACCTGCAATTAACCGTACATACCAGGCAAGTACG 420
Db 13013 GCCTTCCGCGCGGCTGACCTGACCTGCAATTAACCGTACATACCAGGCAAGTACG 13072
Qy 421 TCCAGTCCCGATCCGTTATTTAATCCTTAATAAACTGCGCTTGGCAATCGATTAACGCG 480
Db 13073 TCCAGTCCCGATCCGTTATTTAATCCTTAATAAACTGCGCTTGGCAATCGATTAACGCG 13132
Qy 481 AACGTGACGAGCGGATCTCTGACGAGGAGGAGGCTCAATTGCTGACTTTACCGGCGAT 540
Db 13133 AACGTGACGAGCGGATCTCTGACGAGGAGGAGGCTCAATTGCTGACTTTACCGGCGAT 13192
Qy 541 CGGCAAAACGCGCTTTGCGCAACTGGAACGCGTGTCTTAATTTTCGCAATCAAACTTTGCG 600
Db 13193 CGGCAAAACGCGCTTTGCGCAACTGGAACGCGTGTCTTAATTTTCGCAATCAAACTTTGCG 13252
Qy 601 CTTTAAACGTGAGAAACGAGCAAGAGCTGTTCAATTAACGAGGATTTACATGGAATC 660
Db 13253 CTTTAAACGTGAGAAACGAGCAAGAGCTGTTCAATTAACGAGGATTTACATGGAATC 13312
Qy 661 AAGGTAGGCGCGCAATGCTCATTAACCGGTGCGGTAAAGCTCCGATCAATGCTGAG 720
Db 13313 AAGGTAGGCGCGCAATGCTCATTAACCGGTGCGGTAAAGCTCCGATCAATGCTGAG 13372
Qy 721 GAGATATTTCTCTGCAACAGCAAGGATGCGAGCCGAGTGTGGGAAAGATCAAC 780
Db 13373 GAGATATTTCTCTGCAACAGCAAGGATGCGAGTGTGGGAAAGATCAAC 13432
Qy 781 GATTACACGAGTGAACACCTTGTCTAAGTTTGCAATTAACCGCAATTTTATTTGCTACAA 840
Db 13433 GATTACACGAGTGAACACCTTGTCTAAGTTTGCAATTAACCGCAATTTTATTTGCTACAA 13492
Qy 841 CGACGCGCAGAGGTGCGCGCAGCCGCGCAACCCGTTATTAATTTGATCAAGCAGG 900
Db 13493 CGACGCGCAGAGGTGCGCGCAGCCGCGCAACCCGTTATTAATTTGATCAAGCAGG 13552
Qy 901 TTGACGCGCCATCAACGCAAAACAGCGCTATGTGTGACATTAACCACTTCAAGTCTG 960
Db 13553 TTGACGCGCCATCAACGCAAAACAGCGCTATGTGTGACATTAACCACTTCAAGTCTG 13612
Qy 961 TTTATGCGCGGACACGATTAATCTGGCAAAATCTGGCGCGCACACTGAGCTCAACTGG 1020
Db 13613 TTTATGCGCGGACACGATTAATCTGGCAAAATCTGGCGCGCACACTGAGCTCAACTGG 13672
Qy 1021 AAGCTTCCGGGTAGCGCGGATTAACGCGCGCGGATGTAACCTGATTTGAACGCTGG 1080
Db 13673 AAGCTTCCGGGTAGCGCGGATTAACGCGCGCGGATGTAACCTGATTTGAACGCTGG 13732
Qy 1081 CGTGGCTAAGCATTAACAGCCAGTGTGATTCAGGTTTGGCTGATCTTCAGACTTTACAG 1140

```

```

Db 13733 CGTGGCTAAGCATTAACAGCCAGTGAATTCAGGTTTGGCTGTCTTCCAGACTTTACAG 13792
Qy 1141 CAGATGCTGATTAACGCGCTGCTCAATTAATGCGCGCCGAGAGGTGAACCTGACC 1200
Db 13793 CAGATGCTGATTAACGCGCTGCTCAATTAATGCGCGCCGAGAGGTGAACCTGACC 13852
Qy 1201 CTGGCAGAGTGTGAAGACGAAATGCGCAGGCGATGTGTTGTTGGCAGGTTTACGCA 1260
Db 13853 CTGGCAGAGTGTGAAGACGAAATGCGCAGGCGATGTGTTGTTGGCAGGTTTACGCA 13912
Qy 1261 ATCGTGAATGAAGACGATATCCCGCTTGCAGTTTGTA 1299
Db 13913 ATCGTGAATGAAGACGATATCCCGCTTGCAGTTTGTA 13951

RESULT 6
AAC68294
ID AAC68294 standard; DNA; 20623 BP.
XX
AC AAC68294;
XX
DT 15-SEP-2003 (revised)
DT 20-FEB-2001 (first entry)
XX
DE Lama2/APPA plasmid coding sequence.
XX
KW Transgenic animal; salivary protein; phytase; phosphorus; animal growth;
XX environmental pollution; pig; ds.
XX
OS Mus musculus.
OS Escherichia coli.
XX Chimeric.
XX
PN WO200064247-A1.
XX
PD 02-NOV-2000.
XX
PF 20-APR-2000; 2000WO-CA000430.
XX
PR 23-APR-1999; 99US-0130508P.
XX
PA (UYGU-) UNIV GUELPH.
XX
PI Forsberg CW, Golovan S, Phillips JP;
XX
DR WPI: 2000-687245/67.
XX P-PSDB; AAB36257.
XX
PT Transgenic non-human animal for gastrointestinal tract specific
PT expression of a protein, preferably phytase, comprises a nucleic acid
PT sequence including a heterologous transgene construct encoding the
PT protein.
XX
XX Claim 56; Fig 5; 152pp; English.
XX
CC The present invention provides transgenic animals which produce desired
CC proteins, in this case pigs which expresses phytase in the salivary
CC gland. Low phytase production levels result in phytate in the diet being
CC excreted and causing phosphorus contamination in water, as well as
CC reducing the growth of animals. The invention provides a number of
CC transgenes containing the E. coli APPA phytase coding sequence. (Updated
CC on 15-SEP-2003 to standardise OS field)
XX
SQ Sequence 20623 BP; 5449 A; 4847 C; 4902 G; 5424 T; 0 U; 1 Other;

Query Match 100.0%; Score 1299; DB 3; Length 20623;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGAAAGCATCTTAATCCATTTTATCTCTTCGATTCCTGTTAACCCCGCAATCTGCA 60
Db 12653 ATGAAAGCATCTTAATCCATTTTATCTCTTCGATTCCTGTTAACCCCGCAATCTGCA 12712

```

```

QY 61 TTCGCTCAGAGTGAAGCCGCGAGCTGAAAGCTGAAAGTGTGTGATTTGTCAGTGTGAT 120
DB 12713 TTCGCTCAGAGTGAAGCCGCGAGCTGAAAGTGTGTGATTTGTCAGTGTGAT 12772
QY 121 GTGCGGTCTCAACCAAGGCGACGCAACTGATGTCAGATGTCAACCCGACGACGATGGCCA 180
DB 12773 GTGCGGTCTCAACCAAGGCGACGCAACTGATGTCAGATGTCAACCCGACGATGGCCA 12832
QY 181 ACCGTGGCCGGTAAAACTGGGTTGGTGTGACACCGCGCGTGTGTGAGCTTAATCGCTTATCTC 240
DB 12833 ACCGTGGCCGGTAAAACTGGGTTGGTGTGACACCGCGCGTGTGTGAGCTTAATCGCTTATCTC 12892
QY 241 GGAATTAACCAAGCCAGCGTCTGTGTAGCCGACGGAATGCTGTGGCGAAAAAGGCGTCCCG 300
DB 12893 GGAATTAACCAAGCCAGCGTCTGTGTAGCCGACGGAATGCTGTGGCGAAAAAGGCGTCCCG 12952
QY 301 CAGTCTGGTCAAGGTCCGATTAATTTGCTGATGTGACGAGCGGTACCGGTAAACAGCGGAA 360
DB 12953 CAGTCTGGTCAAGGTCCGATTAATTTGCTGATGTGACGAGCGGTACCGGTAAACAGCGGAA 13012
QY 361 GCGTTCGCGCGCGGCGTGGACCTGACTGTGCAATACCGTACATACCCAGGAGATACG 420
DB 13013 GCGTTCGCGCGCGGCGTGGACCTGACTGTGCAATACCGTACATACCCAGGAGATACG 13072
QY 421 TCCAGTCCCGATCCGTTATTTAATTCCTTAAAACTGGCGTTGGCCACTGTGATACGCG 480
DB 13073 TCCAGTCCCGATCCGTTATTTAATTCCTTAAAACTGGCGTTGGCCACTGTGATACGCG 13132
QY 481 AAGGTATCTGACGCGATCTCAGACAGGGGAGAGGGTCAATTTGCTGACTTTACCGGGCAT 540
DB 13133 AAGGTATCTGACGCGATCTCAGACAGGGGAGAGGGTCAATTTGCTGACTTTACCGGGCAT 13192
QY 541 CGGCAAAACGCGCTTTTCGCAACTGGAACGCGGTGCTTAATTTTCGCAATCAAACTTGTGC 600
DB 13193 CGGCAAAACGCGCTTTTCGCAACTGGAACGCGGTGCTTAATTTTCGCAATCAAACTTGTGC 13252
QY 601 CTTAAACGTAGAAACAGACGAAAGCTTTTCAATTAACGACGATTAACCATCGGAATC 660
DB 13253 CTTAAACGTAGAAACAGACGAAAGCTTTTCAATTAACGACGATTAACCATCGGAATC 13312
QY 661 AAGGTGAGCGCGCAATGTCTCATTAACGGTGTGTAAGCTTCGATCAATGCTGACG 720
DB 13313 AAGGTGAGCGCGCAATGTCTCATTAACGGTGTGTAAGCTTCGATCAATGCTGACG 13372
QY 721 GAGATATTTCTCTGCAACAGCAACAGGAAATGCGGAGCGGCGTGGGAGGATCAC 780
DB 13373 GAGATATTTCTCTGCAACAGCAACAGGAAATGCGGAGCGGCGTGGGAGGATCAC 13432
QY 781 GATTCAACCAAGTGAACACCTTCTTAAGTTTGATTAACGGCAATTTTATTTGCTACAA 840
DB 13433 GATTCAACCAAGTGAACACCTTCTTAAGTTTGATTAACGGCAATTTTATTTGCTACAA 13492
QY 841 CGACGCGCAGAGGTGGCCCGGAGCGCGCACCCCGTTATTAAGTTTGATCAAGACGCG 900
DB 13493 CGACGCGCAGAGGTGGCCCGGAGCGCGCACCCCGTTATTAAGTTTGATCAAGACGCG 13552
QY 901 TTGACGCCCATCCACCGCAAAAACAGCGGTATGTGTGACATTAACCACTTCAGTGTG 960
DB 13553 TTGACGCCCATCCACCGCAAAAACAGCGGTATGTGTGACATTAACCACTTCAGTGTG 13612
QY 961 TTTATCGCGCGACAGATTAATCTGTGCAATCTCGCGCGCGACCTGAGACTCAACTGG 1020
DB 13613 TTTATCGCGCGACAGATTAATCTGTGCAATCTCGCGCGCGACCTGAGACTCAACTGG 13672
QY 1021 ACGCTTCCCGGTGACCGCGGATAACAGCGCGCAGGTGTGAAGTGTGTTGAACGCTGG 1080
DB 13673 ACGCTTCCCGGTGACCGCGGATAACAGCGCGCAGGTGTGAAGTGTGTTGAACGCTGG 13732
QY 1081 CGTGGGCTTAAGCATTAACAGCAGTGAATTCAGGTTTGTGCTGTCTTCAGACTTTTACAG 1140
DB 13733 CGTGGGCTTAAGCATTAACAGCAGTGAATTCAGGTTTGTGCTGTCTTCAGACTTTTACAG 13792
QY 1141 CAGATGCGGTATAAAACGCGCGCTGTCAATTAATACGCGCGCGAGAGAGTGAACCTGACC 1200

```

```

DB 13793 CAGATGCGGTATAAAACGCGCGCTGTCAATTAATACGCGCGCGAGAGGTGAACCTGACC 13852
QY 1201 CTGGCAGATGTGAAGAGCGAAATGCGACGGGCACTGTCTTGGCCAGCTTTACGGCAA 1260
DB 13853 CTGGCAGATGTGAAGAGCGAAATGCGACGGGCACTGTCTTGGCCAGCTTTTACGGCAA 13912
QY 1261 ATCGTAATGAAGACGACATACCGCTTGCAGTTGTAA 1299
DB 13913 ATCGTAATGAAGACGACATACCGCTTGCAGTTGTAA 13951

RESULT 7
ABK12514
ID ABK12514 standard; DNA; 1299 BP.
XX
AC ABK12514;
XX
DT 05-JUN-2002 (first entry)
XX
DE DNA encoding phytase associated protein.
XX
KW Phytase; gene; ds.
XX
OS Unidentified.
XX
FH Key Location/Qualifiers
FT 1..1299
FT CDS /product= "Phytase associated protein"
XX
PN KR9086028-A.
XX
PD 15-DEC-1999.
XX
PF 25-MAY-1998; 98KR-00018810.
XX
PR 25-MAY-1998; 98KR-00018810.
XX
PA (WOOL-) WOOLIN CO LTD.
XX
PI Bae HD, Forceburgh CW, Goloben S, Cheng KU;
XX
DR WPI; 2000-645078/62.
XX
DR P-PSDB; AAU77775.
XX
PT Novel phytase gene, recombinant phytase and usage thereof.
XX
PS Claim 1; Fig 2; 10pp; Korean.
XX
CC The invention relates to a novel phytase gene, a recombinant phytase gene
CC and their uses. This sequence encodes a phytase associated protein,
CC described in the invention
XX
SQ Sequence 1299 BP; 317 A; 345 C; 355 G; 282 T; 0 U; 0 Other;

Query Match 99.6%; Score 1294.2; DB 3; Length 1299;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1296; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGAAGCCATCTTAATCCGATTTTATCTTCTGATTCGTTAACCCGCAATCTGCA 60
DB 1 ATGAAGCCATCTTAATCCGATTTTATCTTCTGATTCGTTAACCCGCAATCTGCA 60
QY 61 TTGCTCAGAGTGAAGCCGAGCTGAAGCTGAAAGTGTGATTTGTCACTGCTGATGT 120
DB 61 TTGCTCAGAGTGAAGCCGAGCTGAAGCTGAAAGTGTGATTTGTCACTGCTGATGT 120
QY 121 GTGCGTCTTCAACCAAGGCGACGCAACTGATGACGATGTCAACCCAGACGATGGCCA 180
DB 121 GTGCGTCTTCAACCAAGGCGACGCAACTGATGACGATGTCAACCCAGACGATGGCCA 180
QY 121 GTGCGTCTTCAACCAAGGCGACGCAACTGATGACGATGTCAACCCAGACGATGGCCA 180
DB 121 GTGCGTCTTCAACCAAGGCGACGCAACTGATGACGATGTCAACCCAGACGATGGCCA 180
QY 181 ACTTGGCGGTTAAACTGGGTTGGCTGACACCGCGCGTGTGATTAATCGCTTATCTC 240

```

Db 181 ACCCTGCGCGTAAACTGGGTTGGCTGACACCGCGCGGTGGAGCTAATGCCCTATCTC 240
QY 241 GGACATTAACCAAGCCAGCGTCTGGTAGCCGACGATTTGTCGCAAAAAAGGCTCCCG 300
Db 241 GGACATTAACCAAGCCAGCGTCTGGTAGCCGACGATTTGTCGCAAAAAAGGCTCCCG 300
QY 301 CAGTCTGGTCAAGTGGCCATTTATTTGCTGATGTCGACGAGCCGTAACCCGTAACAGGCGAA 360
Db 301 CAGTCTGGTCAAGTGGCCATTTATTTGCTGATGTCGACGAGCCGTAACCCGTAACAGGCGAA 360
QY 361 GCTTCGCGCGCGGCTGCGACCTGACTGCAATACCGTACATACCAGGACAGATACG 420
Db 361 GCTTCGCGCGCGGCTGCGACCTGACTGCAATACCGTACATACCAGGACAGATACG 420
QY 421 TCCAGTCCCGGATCCGTTATTTAATCTCTAATAAATGCGGTTTGCCCACTGATTAACGCG 480
Db 421 TCCAGTCCCGGATCCGTTATTTAATCTCTAATAAATGCGGTTTGCCCACTGATTAACGCG 480
QY 481 AACGTGACTGACGCGCATCTCTACGACAGGACGAGAGGCTCAATTGCTGACTTTACCGGCACT 540
Db 481 AACGTGACTGACGCGCATCTCTACGACAGGACGAGAGGCTCAATTGCTGACTTTACCGGCACT 540
QY 541 CGGCAAAACGCGCTTTGCGCAACTGGAACGCGGTCTTAATTTTCCGCAATCAAACTTGTGC 600
Db 541 CGGCAAAACGCGCTTTGCGCAACTGGAACGCGGTCTTAATTTTCCGCAATCAAACTTGTGC 600
QY 601 CTTAAACGTGAGAAACAGAACGAAAGCTGTCTTAACGACGAGCATTAACCATCGGAACCTC 660
Db 601 CTTAAACGTGAGAAACAGAACGAAAGCTGTCTTAACGACGAGCATTAACCATCGGAACCTC 660
QY 661 AAGGTAAAGGCGCGCAATGCTCTCATTAACCGGCGGTAAAGCTGSCATCAATGCTGAGC 720
Db 661 AAGGTAAAGGCGCGCAATGCTCTCATTAACCGGCGGTAAAGCTGSCATCAATGCTGAGC 720
QY 721 GAGATATTTCTCCTGCAACAGCAACGAGAAATGCCGAGCCGCGGTGGGAAAGATCAC 780
Db 721 GAGATATTTCTCCTGCAACAGCAACGAGAAATGCCGAGCCGCGGTGGGAAAGATCAC 780
QY 781 GATTCAACCAAGTGAACACCTTGCTAAGTTGCTAATACCGCAATTTATTTGCTACAA 840
Db 781 GATTCAACCAAGTGAACACCTTGCTAAGTTGCTAATACCGCAATTTATTTGCTACAA 840
QY 841 CGACGCGCAGAGGTTCGCGGACGCGGCAACCCGTTATTAATTTGATCAAGACGCG 900
Db 841 CGACGCGCAGAGGTTCGCGGACGCGGCAACCCGTTATTAATTTGATCAAGACGCG 900
QY 901 TTGACGCCCATCATCACCGCAAAACAGGCGTATGCTGACATTAACCACTTCAGTCTG 960
Db 901 TTGACGCCCATCATCACCGCAAAACAGGCGTATGCTGACATTAACCACTTCAGTCTG 960
QY 961 TTTATGCGCGGACACGATTAATCTTGCGCAATCTGCGCGCGCACTGAGCTCAACTG 1020
Db 961 TTTATGCGCGGACACGATTAATCTTGCGCAATCTGCGCGCGCACTGAGCTCAACTG 1020
QY 1021 AAGCTTCCCGGTAGCGGATTAACGCGCGCGAGGTGTAATGCGTTGTAACCTG 1080
Db 1021 AAGCTTCCCGGTAGCGGATTAACGCGCGCGAGGTGTAATGCGTTGTAACCTG 1080
QY 1081 CGTGGGTAAAGCAATACAGCCAGTGGATTCAAGTTTCCGTTCCAGACTTTACAG 1140
Db 1081 CGTGGGTAAAGCAATACAGCCAGTGGATTCAAGTTTCCGTTCCAGACTTTACAG 1140
QY 1141 CAGATGCGTATTAACAGCGCGCTGCTATTAATACCGCGCGGAGAGGTGAACGAC 1200
Db 1141 CAGATGCGTATTAACAGCGCGCTGCTATTAATACCGCGCGGAGAGGTGAACGAC 1200
QY 1201 CTGGCAGGATGTAAAGGCAAAATGCGCAGGAGATGTTGCTTGGAGGTTTACGCA 1260
Db 1201 CTGGCAGGATGTAAAGGCAAAATGCGCAGGAGATGTTGCTTGGAGGTTTACGCA 1260
QY 1261 ATCGTAATGAGACGCAATACCGCTTGGCAGTTTGTAA 1299

Db 1261 ATCGTAATGAGACGCAATACCGCTTGGCAGTTTGTAA 1299
RESULT 8
ACAI9297
ID ACAI9297 standard; DNA; 1299 BP.
XX
AC ACAI9297;
XX
DT 19-JUN-2003 (first entry)
XX
DE Prokaryotic essential gene #954.
XX
KW Antisense; ds; prokaryotic essential gene; cell proliferation;
XX drug design; gene.
XX
OS Escherichia coli.
XX
PN WO20027183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
XX 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-034923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR MPI; 2003-029926/02.
XX
PT P-PSDB; ABU15427.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 14; SEQ ID NO 7167; 1766bp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound an antibiotic; (10) profiling a
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in

electronic format directly from WIFO at
ftp.wifo.int/pub/published_pcr_sequences

Sequence 1299 BP; 317 A; 345 C; 355 G; 282 T; 0 U; 0 Other;

Query Match 99.6%; Score 1294.2; DB 8; Length 1299;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1296; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 ATGAAAGCATCTTAATCCCATTTTATCTTCTGATTCCTGTAACCCCGCAATCTGCA 60
1 ATGAAAGCATCTTAATCCCATTTTATCTTCTGATTCCTGTAACCCCGCAATCTGCA 60
61 TTGCTCAGAGTGAAGCCGAGCTGAAAGTGTGTGATTTGTCACTGCTCATGCT 120
61 TTGCTCAGAGTGAAGCCGAGCTGAAAGTGTGTGATTTGTCACTGCTCATGCT 120
121 GTGCGCTCCCAACCAAGGCAAGCACTGATGCAAGATGTCAACCCGAGCATGAGCA 180
121 GTGCGCTCCCAACCAAGGCAAGCACTGATGCAAGATGTCAACCCGAGCATGAGCA 180
181 ACCTGGCCGGTAAATCTGGGTTGGCTGACACCGCGGTTGGTGAAGTAAATGCTATCTC 240
181 ACCTGGCCGGTAAATCTGGGTTGGCTGACACCGCGGTTGGTGAAGTAAATGCTATCTC 240
241 GGAATTACCAACGCGAGCTGTGTAGCCGAGTGTGCGGAAAGGCTGCGCG 300
241 GGAATTACCAACGCGAGCTGTGTAGCCGAGTGTGCGGAAAGGCTGCGCG 300
301 CAGTCTGTGATAGTCCGATTAATTTGCTGATGTCAGAGGTACCCGTAACAGAGGGA 360
301 CAGTCTGTGATAGTCCGATTAATTTGCTGATGTCAGAGGTACCCGTAACAGAGGGA 360
361 GCCTTGCCGCGCGGAGCTGACCTGACATAAATCCGTAACCCGAGCATGAC 420
361 GCCTTGCCGCGCGGAGCTGACCTGACATAAATCCGTAACCCGAGCATGAC 420
421 TCCAGTCCCGATCGTTATTTAAATCTCTTAAATCTGGCTTGGCAATGATTAACGG 480
421 TCCAGTCCCGATCGTTATTTAAATCTCTTAAATCTGGCTTGGCAATGATTAACGG 480
481 AAGTGAATGACCGGATCTCTCAGAGGCAAGGAGTCAATGCTGATTTACCGGCAAT 540
481 AAGTGAATGACCGGATCTCTCAGAGGCAAGGAGTCAATGCTGATTTACCGGCAAT 540
541 CGGCAAAACGCGCTTGGCGAATGAAAGGCTTAAATTTCCGCAATCAATCTTGGC 600
541 CGGCAAAACGCGCTTGGCGAATGAAAGGCTTAAATTTCCGCAATCAATCTTGGC 600
601 CTTAAACGTGAGAAACGAGAACGCTGTTCAATTAACGAGCATTAACATCGGAATCTC 660
601 CTTAAACGTGAGAAACGAGAACGCTGTTCAATTAACGAGCATTAACATCGGAATCTC 660
661 AAGTGAAGCCGCAATGTTCTTCAATTAACGAGTGGGTAAGCTCGCATCAATGCTAGC 720
661 AAGTGAAGCCGCAATGTTCTTCAATTAACGAGTGGGTAAGCTCGCATCAATGCTAGC 720
721 GAGATATTTCTCTGCAACAGCAAGAGTGGGTAAGCTCGCATCAATGCTAGC 780
721 GAGATATTTCTCTGCAACAGCAAGAGTGGGTAAGCTCGCATCAATGCTAGC 780
781 GATTCAACAGAGTGAACACCTTGTCTAATGTTGATTAACGAGCATTAATTTGCTCAAA 840
781 GATTCAACAGAGTGAACACCTTGTCTAATGTTGATTAACGAGCATTAATTTGCTCAAA 840
841 CGCAGCGCAAGTGTGGCCGAGCGCGCAACCCGTTATTAAGTTGATCAACAGCG 900
841 CGCAGCGCAAGTGTGGCCGAGCGCGCAACCCGTTATTAAGTTGATCAACAGCG 900
901 TTGAGCGCCCATCAACCGCAAAACAGGCGATGATGATTAACCATTTCACTTCACTGCTG 960
901 TTGAGCGCCCATCAACCGCAAAACAGGCGATGATGATTAACCATTTCACTTCACTGCTG 960

961 TTTATCGCCGAGCAGATACTATCTGCAATCTCGCGGCGCACTGAGCTCAACTGG 1020
961 TTTATCGCCGAGCAGATACTATCTGCAATCTCGCGGCGCACTGAGCTCAACTGG 1020
1021 AGCTTCCCGGCTCAGCCGAGTAAACAGCCGCGAGTGTGAATCTGTTTGAACGCTGG 1080
1021 AGCTTCCCGGCTCAGCCGAGTAAACAGCCGCGAGTGTGAATCTGTTTGAACGCTGG 1080
1081 CGTGGCTTAAGGATTAACAGCCAGTGTGATTTGCTGCTGCTTCCAGACTTACAG 1140
1081 CGTGGCTTAAGGATTAACAGCCAGTGTGATTTGCTGCTGCTTCCAGACTTACAG 1140
1141 CAGATGCGTGAATAACGCGCTGTCTATTAATAGCCCGCGAGAGTGAATGAC 1200
1141 CAGATGCGTGAATAACGCGCTGTCTATTAATAGCCCGCGAGAGTGAATGAC 1200
1201 CTGGCAGATGTGAAGAGCAATGCGAGGCAATGTTGTTGCTGAGCTTTACGAA 1260
1201 CTGGCAGATGTGAAGAGCAATGCGAGGCAATGTTGTTGCTGAGCTTTACGAA 1260
1261 ATCGTAATGAAGACAGCATACCCGCTTGCAGTTTGTAA 1299
1261 ATCGTAATGAAGACAGCATACCCGCTTGCAGTTTGTAA 1299

RESULT 9
ADL16137
ID ADL16137 standard; DNA; 1299 BP.
XX
AC ADL16137;
DT 06-MAY-2004 (first entry)
XX
DE Escherichia coli K-12 Appla phytase gene.
XX
XX Appla phytase; modified phytase; K-12; animal feed; feed additive;
KW phosphorus; phytate; myo-inositol hexakisphosphate; food production;
KW wet grain milling; dental care product; bone resorption; osteoporosis;
KW osteopathic; kidney stone; metal removal; waste treatment;
KW cleaning product; rust removal product; gene; ds.
XX
OS Escherichia coli; K-12.
XX
FH Key location/Qualifiers
FT CDS 1..1299
FT /*tag= a
XX /product= "Appla phytase"
PN NO2004015084-A2.
XX
PD 19-FEB-2004.
XX
PP 11-AUG-2003; 2003MO-US025058.
XX
PR 12-AUG-2002; 2002US-0403330P.
XX
PA (GEMV) GENENCOR INT INC.
XX
PI Blattmann BO, Darzins A, Davis JM, Encell LP, Morrison TB;
PI Mreacko GT, Schellenberger V;
XX
DR WPI; 2004-169498/16.
DR P-PSDS; ADL16129.
XX
XX Novel recombinant phytase having modified phytase activity comprising
PT modification of amino acid residues in mature Escherichia coli phytase
PT designated EB18B2, useful as feed additive.
XX
PS Claim 19; Fig 21; 107pp; English.
XX
XX The invention relates to recombinant modified Appla phytases having
CC modified phytase activity. The Appla phytases are derived from a mature
CC Escherichia coli Appla phytase designated EB18B2 (see ADL16120), and

QY	1	ATGGAAGCATCTTAATCCCATCTTTATCTCTTCAATCCGTTAACCCGCGAATCTGCA	60
Db	1	ATGGAAGCATCTTAATCCCATCTTTATCTCTTCAATCCGTTAACCCGCGAATCTGCA	60
QY	61	TTGCGCTCAGAGTGAGCCGAGCTGAAAGCTGGAAGTGTGGTGAATGTCACTCGTCAATGGT	120
Db	61	TTGCGCTCAGAGTGAGCCGAGCTGAAAGTGTGGTGAATGTCACTCGTCAATGGT	120
QY	121	GTGCGCGGTCCAAACCAAGGCCAAGCAACTGATGACAGATGTCAACCCCAAGCGATGAGCCA	180
Db	121	GTGCGGTGCTTCAAACCAAGGCCAAGCAACTGATGACAGATGTCAACCCCAAGCGATGAGCCA	180
QY	181	ACCTGGCCGGTAAAACTGGGGTTGGCTGAACAACCGCGCGGTGGAGCTAATGCGCTTATCTC	240
Db	181	ACCTGGCCGGTAAAACTGGGGTTGGCTGAACAACCGCGCGGTGGATAGCTAATGCGCTATCTC	240
QY	241	GGACATTACCAACGCGCAGGCGTCTGTGATGCCGAGATTTGCTGGCGAAAAAGGCGTGGCCG	300
Db	241	GGACATTACCAACGCGCGGTCTGTGATGCCGAGATTTGCTGGCGAAAAAGGCGTGGCCG	300
QY	301	CAGTCTGATCAGGTGCGCATTTATTTGCTGATGTGACGACGAGCTAACCCGTAAACAGCGGAA	360
Db	301	CAGTCTGATCAGGTGCGCATTTATTTGCTGATGTGACGAGAGCTAACCCGTAAACAGCGGAA	360
QY	361	GCGTTGGCGCGCGGGGCTGGACCTGACTGTGCAATAACGTAATACCGGAGAGATACG	420
Db	361	GCGTTGGCGCGCGGGGCTGGACCTGACTGTGCAATAACGTAATACCGGAGAGATACG	420
QY	421	TCGAGTCCCGATCCGTTATTTAATTCCTCTTAAAAACTGGCCCTTTGCCAACTGATTAACGCG	480
Db	421	TCGAGTCCCGATCCGTTATTTAATTCCTCTTAAAAACTGGCCTTTGCCAACTGATTAACGCG	480
QY	481	AACGTACTGACGCGATCTCAGCAGGGCAGAGAGGTCAATTGTGTGACTTTACCGGGCAT	540
Db	481	AACGTACTGACCGCATCTCTCAGCAGGGCAGAGAGGTCAATTGTGTGACTTTACCGGGCAT	540

FT	188.1486
FT	188.1486
CDS	188.1486

```

...
FTT
...
/*tag= a

```


/product="E. coli appa phytase protein"

WO200190333-A2.

29-NOV-2001.

24-MAY-2001; 2001WO-US017118.

25-MAY-2000; 2000US-00580515.

(DIVE-) DIVERSA CORP.

Short JM, Kretz KA, Gray KA, Barton NR, Garrett JB, O'donoghue E;

MPI; 2002-083108/11.

P-PSDB; AAE15807.

New bacterial phytase for e.g. improving the nutritional value of phytate-containing foodstuffs and subsequently improving the growth performance of an organism that consumes it, or in treating animal digestive systems.

Claim 42; Fig 7; 170pp; English.

The patent discloses recombinant bacterial phytase from *Escherichia coli* K12 appa phytase. The enzyme has phytase activity and improved thermal tolerance when compared with wild-type phytase. It has improved protease stability at low pH. The recombinant phytase is useful for improving the nutritional value of phytate-containing foodstuffs and subsequently improving the growth performance of an organism that consumes it, in treating animal digestive systems, in feed treatment processes and for in vitro purposes related to research, discovery and development. They are also used for generating recombinant digestive system life forms, for producing or manufacturing alcoholic and non-alcoholic drinks based on the use of moulds, grains and/or plants, in biopulping and bio-bleaching where a reduction in the use of environmentally harmful chemicals that are traditionally used in the pulp and paper industry is desired and in the reduction or possible elimination of the need for mineral supplements, enzymes or therapeutic drugs for animals from the daily feed. CC thus increasing the amount calories and nutrients present in the feed. CC The present sequence is a DNA encoding *E. coli* appa phytase wild type protein

Sequence 1901 BP; 474 A; 499 C; 499 G; 428 T; 0 U; 1 Other;

Query Match 99.6%; Score 1293.2; DB 6; Length 1901;

Best Local Similarity 99.7%; Pred. No. 0; Matches 1295; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

1 ATGAAAGCCATCTTAATCCATTTTATCTCTTCTGATTCGTTAACCCGCAATCTGCA 60
 188 ATGAAAGCCATCTTAATCCATTTTATCTCTTCTGATTCGTTAACCCGCAATCTGCA 247
 61 TTGGCTCAGAGTACCGGAGCTGAAAGTGTGATGTTGTCAGTCCGTCATGCT 120
 248 TTGGCTCAGAGTACCGGAGCTGAAAGTGTGATGTTGTCAGTCCGTCATGCT 307
 121 GTGGCTCAGAGTACCGGAGCTGAAAGTGTGATGTTGTCAGTCCGTCATGCT 180
 308 GTGGCTCAGAGTACCGGAGCTGAAAGTGTGATGTTGTCAGTCCGTCATGCT 367
 181 ACCTGGCCGATTAATCTGGTGTGCTGACACCGCGCGTGTGATGCTTAATGCTTATCTC 240
 368 ACCTGGCCGATTAATCTGGTGTGCTGACACCGCGCGTGTGATGCTTAATGCTTATCTC 427
 241 GGAATATACCAAGCCGAGCTGTGATGCGGAGTTCGTTGAGGAAAGGGCTGCGCG 300
 428 GGAATATACCAAGCCGAGCTGTGATGCGGAGTTCGTTGAGGAAAGGGCTGCGCG 487
 301 CAGTCTGTGATGCGGATGTTGATGATGTCAGACGAGCGCTAACCCGTTAAACAGGCGAA 360
 488 CAGTCTGTGATGCGGATGTTGATGATGTCAGACGAGCGCTAACCCGTTAAACAGGCGAA 547
 361 GCCTTCCGCGCGGAGCTGAGCACTGATGTCATTAACCGTACATACCAGGCGAGATACG 420

548 GCCTTCCGCGCGGAGCTGAGCACTGATGTCATTAACCGTACATACCAGGCGAGATACG 607
 421 TCCAGTCCGATCCGTTATTTATCTCTTCTGATTCGTTAACCCGCAATCTGATACGCG 480
 608 TCCAGTCCGATCCGTTATTTATCTCTTCTGATTCGTTAACCCGCAATCTGATACGCG 667
 481 AACGTGATGAGCGGATTCCTGACAGGCGAGGAGGTCATTTGCTGACTTACCGGCGAT 540
 668 AACGTGATGAGCGGATTCCTGACAGGCGAGGAGGTCATTTGCTGACTTACCGGCGAT 727
 541 CCGCAAAACGCGCTTTCGCAACTGTAACCGGCTTAAATTTTCCGCAATCAACTTTGTC 600
 728 CCGCAAAACGCGCTTTCGCAACTGTAACCGGCTTAAATTTTCCGCAATCAACTTTGTC 787
 601 CTTAAACGTGAGAAACAGGACGAAAGCTTTCATTAACGAGGATTAACATTCGGAATC 660
 788 CTTAAACGTGAGAAACAGGACGAAAGCTTTCATTAACGAGGATTAACATTCGGAATC 847
 661 AAGGTGAGCGCGGACAAATGTCATTAACCGGCTGTAAGCTGATCAATGTCAGC 720
 848 AAGGTGAGCGCGGACAAATGTCATTAACCGGCTGTAAGCTGATCAATGTCAGC 907
 721 GAGATATTTCTCTGCAACAGACAGGAAATGCGAGCCGCGGTGCGGAAAGATCACC 780
 908 GAGATATTTCTCTGCAACAGACAGGAAATGCGAGCCGCGGTGCGGAAAGATCACC 967
 781 GATTACACCAAGTGAACACCTTGTCAAGTTGTCATTAACGCGCAATTTTATTTGCTACA 840
 968 GATTACACCAAGTGAACACCTTGTCAAGTTGTCATTAACGCGCAATTTTATTTGCTACA 1027
 841 CCGACCCCAAGGTTTCCGCGAGCGCGCACCCGCTTATTAATGATCAAGACAGCG 900
 1028 CCGACCCCAAGGTTTCCGCGAGCGCGCACCCGCTTATTAATGATCAAGACAGCG 1087
 901 TTGACCCCAATCACCAGCAAAACAGGCTATGATGTCATTAACCACTTCAAGTCTG 960
 1088 TTGACCCCAATCACCAGCAAAACAGGCTATGATGTCATTAACCACTTCAAGTCTG 1147
 961 TTTATGCGCGGACAGATTAATCTGCAATCTGCGGCGGACCTGAGCTCAACTG 1020
 1148 TTTATGCGCGGACAGATTAATCTGCAATCTGCGGCGGACCTGAGCTCAACTG 1207
 1021 ACCGTTCCGCTGACCGGATTAACAGCGCGGAGTGTGAATGAGTTTGAACGCTG 1080
 1208 ACCGTTCCGCTGACCGGATTAACAGCGCGGAGTGTGAATGAGTTTGAACGCTG 1267
 1081 CGTGGCTAAAGCATTAACAGCCAGTGAATCAGGTTTCGCTGCTTCCAGACTTTACAG 1140
 1268 CGTGGCTAAAGCATTAACAGCCAGTGAATCAGGTTTCGCTGCTTCCAGACTTTACAG 1327
 1141 CAGATGCTGATTAACAGCCGCTGTGATTAATACGCGCGGAGAGTGAACCTGACC 1200
 1328 CAGATGCTGATTAACAGCCGCTGTGATTAATACGCGCGGAGAGTGAACCTGACC 1387
 1201 CTGGCAGAGTGAAGAGCAAAATGCGCAGGCGCATGTTGCTTGGCAGGTTTACGCA 1260
 1388 CTGGCAGAGTGAAGAGCAAAATGCGCAGGCGCATGTTGCTTGGCAGGTTTACGCA 1447
 1261 ATGCTGAATGAACAGCATATCCGCTGAGCTTGTGA 1289
 1448 ATGCTGAATGAACAGCATATCCGCTGAGCTTGTGA 1486

RESULT 11
 ADA19449
 ID ADA19449 standard; DNA; 1901 BP.
 AC ADA19449;
 XX 20-NOV-2003 (first entry)
 DE E. coli K12 appa gene encoding phytase.

XX Phytase; ds; gene; phytase; appa gene; animal feed; inorganic phosphate;
 KW digestion enhancement; transgenic; thermal tolerance; protease stability.
 XX Escherichia coli; strain K12.
 XX Key location/Qualifiers
 FT CDS 188..1486
 FT /tag=a
 FT /product="Phytase"
 XX US2002136754-A1.
 PD 26-SEP-2002.
 PF 24-MAY-2001; 2001US-00866379.
 XX 13-AUG-1997; 97US-00910798.
 PR 01-MAR-1999; 99US-00259214.
 PR 13-APR-1999; 99US-00291931.
 PR 25-MAY-1999; 99US-00318528.
 PR 25-MAY-2000; 2000US-00580515.
 PA (SHOR/) SHORT J M.
 PA (KRETT/) KRETTZ K A.
 PA (GRAY/) GRAY K A.
 PA (BART/) BARTON N R.
 PA (GARRE/) GARRETT J B.
 PA (DONC/) O' DONOGHUE E.
 PA (MATHU/) MATHUR E J.
 PI Short JM, Kretz KA, Gray KA, Barton NR, Garrett JB;
 PI O' Donoghue E, Mathur EJ;
 DR MPI; 2003-040002/03.
 XX P-PADB; ADA19450.
 PT Isolated Escherichia coli polynucleotide encoding a modified phytase
 PT enzyme, useful in the production of animal feed, for improving the
 PT nutritional value of phytase-containing feedstuff and for enhancing
 PT digestion in humans and animals.
 XX Claim 42; Fig 7; 62pp; English.
 PS The invention relates to an isolated Escherichia coli polynucleotide
 CC encoding a phytase enzyme appearing as ADA19450 and having amino acids
 CC modification from W68E, Q84W, A95P, K97C, S168E, R180Y, N226C or Y277D.
 CC Also included the E. coli appa gene ADA19449 (or an oligonucleotide
 CC derived from it) or its mutant sequence ADA19452, expression vectors,
 CC host cells, a method of improving nutritional value of a phytase-
 CC containing feedstuff by contacting the phytase-containing feedstuff with
 CC a pure phytase enzyme mutated phytase (where the phytase enzyme catalyzes
 CC the liberation of inorganic phosphate from the phytate in the phytate-
 CC containing feedstuff), a method to produce an animal feed containing a
 CC microbial phytase (comprising culturing the plant cell, plant part or
 CC plant under conditions where the nucleotide sequence is expressed and
 CC converting the plant cells, plants or plant into a composition for animal
 CC feed), a feed composition for animals (comprising the plant seeds, plant
 CC cells, plant parts or plants in admixture with a phytase-containing
 CC feedstuff), a method to treat a human or an animal able to benefit from
 CC digestive enhancement by the activity of an exogenous phytase enzymes
 CC comprising administering to the human or animal the plant seed, plant
 CC cell, plant parts or plants of a transgenic plant which is modified to
 CC contain an expression system which expresses a nucleotide sequence
 CC encoding a phytase enzyme, a transgenic non-human organism whose genome
 CC comprising a heterologous nucleic acid sequence encoding a polypeptide
 CC having phytase activity. The phytase enzyme is useful for improving the
 CC nutritional value of phytase-containing feedstuff, in the production of
 CC animal feed and for enhancing digestion in humans and animals. The
 CC invented method improves thermal tolerance and protease stability. It
 CC also improves the feeding value of phytate rich ingredients. The present
 CC sequence represents the E. coli K12 appa gene encoding wild-type phytase.
 XX

Sequence 1901 BP; 474 A; 499 C; 499 G; 428 T; 0 U; 1 Other;
 Query Match 99.6%; Score 1293.2; DB 8; Length 1901;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 1295; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 1 ATGAAAGCCATCTTAATCCATTTTATCTTCTGATTCGTTAACCCGCAATCTGCA 60
 DB 188 ATGAAAGCCATCTTAATCCATTTTATCTTCTGATTCGTTAACCCGCAATCTGCA 247
 QY 61 TTGCTCAGAGTAGAGCCGAGAGCTGAAGCTGGAAGTGGTGAATTTGATGTCATGCT 120
 DB 248 TTGCTCAGAGTAGAGCCGAGAGCTGAAGCTGGAAGTGGTGAATTTGATGTCATGCT 307
 QY 121 GTGCTGTCTCAACCAAGCCCAAGCACTGATGAGAGATGTCACCCGAGAGCATAGGCA 180
 DB 308 GTGCTGTCTCAACCAAGCCCAAGCACTGATGAGAGATGTCACCCGAGAGCATAGGCA 367
 QY 181 ACCTGCGCGGTAAACTGGTGTGCTGACACCGCGCGGTGCTGAGCTAATCGCTATCTC 240
 DB 368 ACCTGCGCGGTAAACTGGTGTGCTGACACCGCGCGGTGCTGAGCTAATCGCTATCTC 427
 QY 241 GGCATTAACCAAGCCGAGAGCTGCTGAGAGCGGATTTGCTGCGGCAAAAAGGCGTCCG 300
 DB 428 GGCATTAACCAAGCCGAGAGCTGCTGAGAGCGGATTTGCTGCGGCAAAAAGGCGTCCG 487
 QY 301 CAGTCTGCTCAGAGTTCGAGATTATGCTGATGTCAGAGCGGTAACCGTAAACAGGCGAA 360
 DB 488 CAGTCTGCTCAGAGTTCGAGATTATGCTGATGTCAGAGCGGTAACCGTAAACAGGCGAA 547
 QY 361 GCTTTCGCGCGCGGCTGAGAGCTGAGCTGTCGATTAACCGTAAACCGTAAACCGAGAGATACG 420
 DB 548 GCTTTCGCGCGCGGCTGAGAGCTGAGCTGTCGATTAACCGTAAACCGTAAACCGAGAGATACG 607
 QY 421 TCCAGTCCGATCCGTTATTTATATCTCTTAAACAGTCCGTTTGGCACTGATTAACGCG 480
 DB 608 TCCAGTCCGATCCGTTATTTATATCTCTTAAACAGTCCGTTTGGCACTGATTAACGCG 667
 QY 481 AACGAGCTGAGAGGATCTGAGAGGAGGAGGAGGATTAATGCTGATTAACGAGGAT 540
 DB 668 AACGAGCTGAGAGGATCTGAGAGGAGGAGGAGGATTAATGCTGATTAACGAGGAT 727
 QY 541 CGGCAACGCGGCTTTCGGAACCTGGAACGCGGCTTAAATTTTCGCAATCAAACTTGTGC 600
 DB 728 CGGCAACGCGGCTTTCGGAACCTGGAACGCGGCTTAAATTTTCGCAATCAAACTTGTGC 787
 QY 601 CTTAAACGTGAGAAACAGAGCAAAAGCTGTTATTAACGAGGCTTAATCCATCGGAATCC 660
 DB 788 CTTAAACGTGAGAAACAGAGCAAAAGCTGTTATTAACGAGGCTTAATCCATCGGAATCC 847
 QY 661 AAGGTAGAGGCGGCAATGCTCATTAACCGGTGCGTAAAGCTCGCATCAATGCTGACG 720
 DB 848 AAGGTAGAGGCGGCAATGCTCATTAACCGGTGCGTAAAGCTCGCATCAATGCTGACG 907
 QY 721 GAGATATTTCTCTCTCAACCAAGCAAGGAAATGCCGAGCGGGTGGGAAAGATCAACC 780
 DB 908 GAGATATTTCTCTCTCAACCAAGCAAGGAAATGCCGAGCGGGTGGGAAAGATCAACC 967
 QY 781 GATTACACCAAGTGAACACCTTGTGATTAACGCGCAATTTTATTTGCTACAA 1027
 DB 968 GATTACACCAAGTGAACACCTTGTGATTAACGCGCAATTTTATTTGCTACAA 1027
 QY 841 GCGAGCGCGAGAGTTCGCGAGCGCGCAACCCGTTATTAATGATTTGATCAAGACAGCG 900
 DB 1028 GCGAGCGCGAGAGTTCGCGAGCGCGCAACCCGTTATTAATGATTTGATCAAGACAGCG 1087
 QY 901 TTGAGCGCCCATCAACCGCAAAACAGGCGTATGATGATCAATCAATCAATCAATCAAT 960
 DB 1088 TTGAGCGCCCATCAACCGCAAAACAGGCGTATGATGATCAATCAATCAATCAATCAAT 1147
 QY 961 TTTATGCGCGGACAGATTAATCTGGAATTTCCGCGGCGCATGGAAGCTCAACTGG 1020
 DB 1148 TTTATGCGCGGACAGATTAATCTGGAATTTCCGCGGCGCATGGAAGCTCAACTGG 1207

QY 1021 ACGCTTCCCGGTACCGCGATTAACACGCCGCCAGGTGTGAACCTGTTTGAACCTCG 1080
DB 1208 ACGCTTCCCGGTACCGCGATTAACACGCCGCCAGGTGTGAACCTGTTTGAACCTCG 1267
QY 1081 CCTCGGCTAAGGATTAACGCCAGTGTGATTCAGCTTCTCCAGACTTAAACAG 1140
DB 1268 CCGGCTAAGGATTAACGCCAGTGTGATTCAGCTTCTCCAGACTTAAACAG 1327
QY 1141 CAGATCGGTATTAACGCCCGTGTATTAATACGCCGCCAGAGGTGAACCTGACC 1200
DB 1328 CAGATCGGTATTAACGCCCGTGTATTAATACGCCGCCAGAGGTGAACCTGACC 1387
QY 1201 CTGGCAGATGTGAAGAGCGAAATGCGAGGCGATGTCTGTTGCAAGTTTACGCA 1260
DB 1388 CTGGCAGATGTGAAGAGCGAAATGCGAGGCGATGTCTGTTGCAAGTTTACGCA 1447
QY 1261 ATCGTGAATGAAGCAGCATACCGGCTTGCAAGTTTGA 1299
DB 1448 ATCGTGAATGAAGCAGCATACCGGCTTGCAAGTTTGA 1486

RESULT 12
ADC87744
ID ADC87744 standard; DNA; 1901 BP.
XX
AC ADC87744;
XX
DT 01-JAN-2004 (first entry)
DE DNA encoding Escherichia coli appA phytase.
XX
KM Phytase; food supplement; enzyme delivery matrix; soybean meal;
KM thermotolerance; thermostability; kernel; phytate;
KM myo-inositol-hexaphosphate; inositol; inorganic phosphate;
KM thermotolerant; feed value; digestion; gene; ds; appA.
XX
OS Escherichia coli.
XX
FH Location/Qualifiers
FT 188..1486
FT CDS
FT /*tag= a
FT /product= "appA phytase"
FT /transl_except= (pos:401..403,aa:Arg)

US2003103958-A1.
XX
PN 05-JUN-2003.
XX
PD 24-MAY-2002; 2002US-00156660.
XX
PF 13-AUG-1997; 97US-00910798.
XX
PR 01-MAR-1999; 99US-00259214.
XX
PR 13-APR-1999; 99US-00291931.
XX
PR 25-MAY-1999; 99US-00318528.
XX
PR 25-MAY-2000; 2000US-00580515.
XX
PR 24-MAY-2001; 2001US-00866379.
XX
PA (DIVE-) DIVERSA CORP.
XX
PI Short JM, Kretz K, Gray KA, Barton NR, Garrett JB, O'donoghue E;
PI Mathur BJ;
XX
DR WPI: 2003-787039/74.
XX
DR P-PSDB; ADC87745.
XX
PT New nucleic acid encoding a polypeptide having phytase activity, useful
XX in improving the feeding value of phytate rich ingredients or as an aid
XX in phytate digestion.
XX
PS Example 1, SEQ ID NO 3; 113pp; English.
XX
CC The invention discloses a new isolated or recombinant nucleic acid which

CC encodes a polypeptide having a phytase activity. Also claimed is a
CC nucleic acid probe, an amplification primer sequence pair, an expression
CC cassette comprising the nucleic acid, a vector comprising the nucleic
CC acid, a transgenic non-human animal or plant, or its seed, comprising the
CC nucleic acid, an antisense oligonucleotide, inhibiting the translation of
CC a phytase message in a cell, a heterodimer comprising the polypeptide and
CC a second domain, an array comprising immobilised polypeptide or nucleic
CC acid, a hydridoma comprising an antibody that specifically binds to the
CC polypeptide, a food supplement for an animal, an edible enzyme delivery
CC matrix, an edible pellet comprising a granule edible carrier and the
CC polypeptide, a feed composition, a soybean meal, isolating or identifying
CC the polypeptide, making an anti-phytase antibody, producing a recombinant
CC polypeptide, determining whether a compound binds to the polypeptide,
CC identifying a modulator whole cell engineering of new or modified
CC phenotypes by using real-time metabolic flux analysis, increasing
CC thermotolerance or thermostability of the phytase polypeptide, increasing
CC the resistance of the phytase polypeptide to enzymatic inactivation in a
CC digestive system and processing of corn and sorghum kernels. The phytase
CC activity comprises catalysis of phytate (myo-inositol-hexaphosphate) to
CC inositol and inorganic phosphate and is thermotolerant. The nucleic acid
CC is useful in improving the feeding value of phytate rich ingredients or
CC as an aid in phytate digestion. The sequence presented is the DNA
CC encoding the Escherichia coli appA phytase.
XX
SQ Sequence 1901 BP; 474 A; 499 C; 499 G; 428 T; 0 U; 1 Other;
XX
Query Match 99.6%; Score 1293.2; DB 10; Length 1901;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1295; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 ATGAAGCCATCTTATCCCATTTTATCTCTGTGATTCGTTAAACCCGCAATCTGA 60
DB 188 ATGAAGCCATCTTATCCCATTTTATCTCTGTGATTCGTTAAACCCGCAATCTGA 247
QY 61 TTGCTCAGAGTGAAGCCGAGCTGAAGCTGAAGTGTGTGATTTGTCACTGTATGT 120
DB 248 TTGCTCAGAGTGAAGCCGAGCTGAAGCTGAAGTGTGTGATTTGTCACTGTATGT 307
QY 121 GTGCGTCTCAACCAAGCCACGCAATGTATGAGATTCACCCGAGCGATGGCCA 180
DB 308 GTGCGTCTCAACCAAGCCACGCAATGTATGAGATTCACCCGAGCGATGGCCA 367
QY 181 ACCTGGCCGTTAAACCTGGTGTGCTGACACCGCGCGTGTGATTCGCTATCTC 240
DB 368 ACCTGGCCGTTAAACCTGGTGTGCTGACACCGCGCGTGTGATTCGCTATCTC 427
QY 241 GGAATTATCCCAAGCCAGCTGTGTAGCCGAGATTTGTGGGAAAAAGGCTGCCG 300
DB 428 GGAATTATCCCAAGCCAGCTGTGTAGCCGAGATTTGTGGGAAAAAGGCTGCCG 487
QY 301 CAGCTGTGCAAGTCCGATTAATTGCTGATGTGACGAGCTTACCCGTAAACAGCGGA 360
DB 488 CAGCTGTGCAAGTCCGATTAATTGCTGATGTGACGAGCTTACCCGTAAACAGCGGA 547
QY 361 GCTTGGCCGCGCGGCTGAGCACTGTGCAATACGTTACATACCGAGCGATAGC 420
DB 548 GCTTGGCCGCGCGGCTGAGCACTGTGCAATACGTTACATACCGAGCGATAGC 607
QY 421 TCCAGTCCGATCGTATTTAAATCTCTAAATACGCGCTTTCGCAATGATTAACGG 480
DB 608 TCCAGTCCGATCGTATTTAAATCTCTAAATACGCGCTTTCGCAATGATTAACGG 667
QY 481 AACGTGATCAAGCGATCTTCAAGAGGCGAGAGGCTCAATTGCTGATTTACCGGCGAT 540
DB 668 AACGTGATCAAGCGATCTTCAAGAGGCGAGAGGCTCAATTGCTGATTTACCGGCGAT 727
QY 541 CCGCAAAAGCGCTTTCGCAATGGAACGCGGTCTTAATTTCCGCAATCAATCTTGC 600
DB 728 CCGCAAAAGCGCTTTCGCAATGGAACGCGGTCTTAATTTCCGCAATCAATCTTGC 787
QY 601 CTTAAAGTGAAGAAACAGAGCAAGAGCTGTTCATTTACGAGGATTAACATCGGAATC 660
DB 788 CTTAAAGTGAAGAAACAGAGCAAGAGCTGTTCATTTACGAGGATTAACATCGGAATC 847

```

QY 661 AAGGTAGGCGCCGACAAATGCTCTATTAAACGGTGGCGGTAGACCTGCAATCAATGCTGACG 720
    |||||
Db 848 AAGGTAGGCGCCGACAAATGCTCTATTAAACGGTGGCGGTAGACCTGCAATGCTGACG 907
    |||||
QY 721 GAGATATTTCTCTCTCAACAAGCAGAGGAATGCGGAGCGGGGTGGGGAAGATCAAC 780
    |||||
Db 908 GAGATATTTCTCTCTCAACAAGCAGAGGAATGCGGAGCGGGGTGGGGAAGATCAAC 967
    |||||
QY 781 GATTGACACAGAGTGAACAACCTTGTGTAAGTTGATTAACGCGCAATTTTATTGTGTACA 840
    |||||
Db 968 GATTGACACAGAGTGAACAACCTTGTGTAAGTTGATTAACGCGCAATTTTATTGTGTACA 1027
    |||||
QY 841 GCGACGCCAGAGGTGGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900
    |||||
Db 1028 GCGACGCCAGAGGTGGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1087
    |||||
QY 901 TTGAGCGCCCATCCACCGCAAAAACAGCGGTATGGTGAATTAACCCACTTCAAGTGTG 960
    |||||
Db 1088 TTGAGCGCCCATCCACCGCAAAAACAGCGGTATGGTGAATTAACCCACTTCAAGTGTG 1147
    |||||
QY 961 TTTATCGCGGACACAGATCTAATCTGGAATTTGCGGCGCGCACTGAGCTCAACTGG 1020
    |||||
Db 1148 TTTATCGCGGACACAGATCTAATCTGGAATTTGCGGCGCGCACTGAGCTCAACTGG 1207
    |||||
QY 1021 AGCGTTCCCGGTACAGCGCGATTAACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1080
    |||||
Db 1208 AGCGTTCCCGGTACAGCGCGATTAACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1267
    |||||
QY 1081 CGTGGGCTAAGCGATTAACAGCGCGATTAACAGCGCGCGCGCGCGCGCGCGCGCGCGCG 1140
    |||||
Db 1268 CGTGGGCTAAGCGATTAACAGCGCGATTAACAGCGCGCGCGCGCGCGCGCGCGCGCGCG 1327
    |||||
QY 1141 CAGATCGGTATTAACAGCGCGCTGTCTTAATATACCGCGCGCGCGCGCGCGCGCGCGCG 1200
    |||||
Db 1328 CAGATCGGTATTAACAGCGCGCTGTCTTAATATACCGCGCGCGCGCGCGCGCGCGCGCG 1387
    |||||
QY 1201 CTGGCAGGATGTAAAGGGAATGCGCAGCGCATGTTCTGTTGCGAGGTTTACGAA 1260
    |||||
Db 1388 CTGGCAGGATGTAAAGGGAATGCGCAGCGCATGTTCTGTTGCGAGGTTTACGAA 1447
    |||||
QY 1261 ATCGTGAATGAACGAGCATACCGCGCTTGCAAGTTGTAA 1299
    |||||
Db 1448 ATCGTGAATGAACGAGCATACCGCGCTTGCAAGTTGTAA 1486
    |||||

RESULT 13
AD050297
ID AD050297 standard; DNA; 1901 BP.
XX
AC AD050297;
XX
DT 29-JUL-2004 (first entry)
XX
DE Escherichia coli K12 apA phytase DNA.
XX
KM ApA phytase; bacteri; thermal tolerance; protease stability; foodstuff;
KW phytate; animal feed; fish feed; dough; baking; gene; ds.
XX
OS Escherichia coli; K12.
XX
FH Key Location/Qualifiers
FT CDS 188..1486
FT /tag= a
FT /product= "ApA phytase"
FT 401..403
FT /tag= b
FT /note= "Encodes Arg"
XX
XX US2004091968-A1.
XX
XX 13-MAY-2004.
XX

```

```

PF 20-JUN-2003; 2003US-00601319.
XX
XX 13-AUG-1997; 97US-00910798.
PR 01-MAR-1999; 99US-00259214.
PR 13-APR-1999; 99US-00291931.
PR 25-MAY-1999; 99US-00318528.
PR 25-MAY-2000; 2000US-00580515.
PR 24-MAY-2001; 2001US-00866379.
XX
XX (SHOR/) SHORT J M.
PA (KRET/) KRETZ K.
PA (GRAY/) GRAY K A.
PA (BART/) BARTON N R.
PA (GAR/) GARRETT J B.
PA (ODON/) O'DONOGHUE E.
PA (MATH/) MATHER E J.
XX
XX Short JM, Kretz K, Gray KA, Barton NR, Garrett JB, O'donoghue E;
PI Mather EJ;
PI
XX
DR MPI: 2004-374952/35.
DR P-PSDB: AD050298.
XX
XX Producing phytase, involves providing nucleic acid derived from bacteria
PT encoding polypeptide having phytase activity, and expressing nucleic acid
PT in yeast.
XX
PS Disclosure; SEQ ID NO 7; 74dp; English.
XX
XX The invention relates to a method for producing phytase that involves
CC providing a nucleic acid encoding phytase derived from a bacteria, and
CC expressing the nucleic acid in a yeast under conditions that allow
CC expression of the enzyme in the yeast. The invention also relates to
CC modified phytase enzyme which has improved thermal tolerance and protease
CC stability at low pH. The phytase enzyme can be used in foodstuffs to
CC improve the feeding value of phytate rich ingredients, and in diet of
CC numerous animals including mammals, fowls and fishes, commercially
CC significant mammals such as pigs, goats, laboratory rodents, commercially
CC significant avian species such as chicken, ducks, doves, parrot, etc.,
CC commercially farmed fish such as guppy, zebrafish, molly, swordtail,
CC etc., in dough making and baking, in dietary aids for animals. The method
CC provides easy manufacture of the active ingredient loaded biocompatible
CC composition, higher yields and loading efficiency. The phytase
CC incorporated in to the dietary aid is safe for animals. The present
CC sequence is Escherichia coli K12 apA phytase DNA.
XX
XX Sequence 1901 BP; 474 A; 499 C; 499 G; 428 T; 0 U; 1 Other;
SQ
XX
XX Query Match 99.6%; Score 1293.2; DB 12; Length 1901;
XX Best Local Similarity 99.7%; Pred. No. 0;
XX Matches 1295; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
XX
QY 1 ATGAAAGCATCTTAATCCCATTTTATCTCTTGATTCGCTTAACCCCGAATCTGCA 60
    |||||
Db 188 ATGAAAGCATCTTAATCCCATTTTATCTCTTGATTCGCTTAACCCCGAATCTGCA 247
    |||||
QY 61 TTGGCTCAGAGTGAGCGCGGAGCTGAAGCTGGAAGTGTGATTTGTCAGTGTATGCT 120
    |||||
Db 248 TTGGCTCAGAGTGAGCGCGGAGCTGAAGCTGGAAGTGTGATTTGTCAGTGTATGCT 307
    |||||
QY 121 GTGGCTGCTCAACCAAGGCGACGCAACTGATGCAAGATGTCAACCCGACGATGGCCA 180
    |||||
Db 308 GTGGCTGCTCAACCAAGGCGACGCAACTGATGCAAGATGTCAACCCGACGATGGCCA 367
    |||||
QY 181 ACCTGGCGCGTAAACTGGGTTGGCTGACACCGCGCGGTGGAGCTATCCCTATCTC 240
    |||||
Db 368 ACCTGGCGCGTAAACTGGGTTGGCTGACACCGCGCGGTGGAGCTATCCCTATCTC 427
    |||||
QY 241 GGACATTACCAAGCGCAGCGCTGCTGTAAGCGGATTTCTGGCGCAAAAAGGCTGCCG 300
    |||||
Db 428 GGACATTACCAAGCGCAGCGCTGCTGTAAGCGGATTTCTGGCGCAAAAAGGCTGCCG 487
    |||||
QY 301 CAGTCTGTGTCAGGTGCGGATTAATTTGCTGATGTGACGAGCGTACCCTGTAACAGGCGAA 360
    |||||

```

Db	488	CAGTCTGGTCAAGTTCGGGATTTATTTGCTGATGTGCAACGAGCTACCCCTTAACAGGGGAA	547
Oy	361	GCCTTCGCGCGCGGGCTGGCACCTGACTGTGCAATTAACCGTACATACCAGGCGATACG	420
Db	548	GCCTTCGCGCGCGGGCTGGCACCTGACTGTGCAATTAACCGTACATACCAGGCGAGATACG	607
Oy	421	TCCAGTCCCGGATCCGGTTATTTTAATCTCTTAAATACTGGCGTTTGGCAACTGGATTAACGC	480
Db	608	TCCAGTCCCGGATCCGGTTATTTTAATCTCTTAAATACTGGCGTTTGGCAACTGGATTAACGC	667
Oy	481	AACGTGACTGACGCGATCTCAGCAGGAGCAGAGGGTCAATTCGTGACTTTACCGGGCAT	540
Db	668	AACGTGACTGACGCGATCTCAGCAGGAGCAGAGGGTCAATTCGTGACTTTACCGGGCAT	727
Oy	541	CGGCAAAACGGCGTTTCGGGAACTGGAAACGGGTGCTTAATTTTCCGCAATCAAACTTGTGC	600
Db	728	CGGCAAAACGGCGTTTCGGGAACTGGAAACGGGTGCTTAATTTTCCGCAATCAAACTTGTGC	787
Oy	601	CTTAAACGTGAGAAACAGAGCAAAAGCTGTTCATTTACGACGACTTAACATCGGAATC	660
Db	788	CTTAAACGTGAGAAACAGAGCAAAAGCTGTTCATTTACGACGACTTAACATCGGAATC	847
Oy	661	AAGGTAGCGCCGACAATGTCTCATTTAACGGTGGGTAAAGCTTCGCATCACTGTGAC	720
Db	848	AAGGTAGCGCCGACAATGTCTCATTTAACGGTGGGTAAAGCTTCGCATCACTGTGAC	907
Oy	721	GAGATATTCTCTGCAACAGACACAGGAAATGCGGAGCCGGGTGGGGAAGATCAAC	780
Db	908	GAGATATTCTCTGCAACAGACACAGGAAATGCGGAGCCGGGTGGGGAAGATCAAC	967
Oy	781	GATTCAACAACAAGTGAACAACCTTGCTAAAGTTTGCAATTAACGGGCAATTTTATTTGCTCAA	840
Db	968	GATTCAACAACAAGTGAACAACCTTGCTAAAGTTTGCAATTAACGGGCAATTTTATTTGCTCAA	1027
Oy	841	CGCAGCCGCAAGGTGGCCCGGACGCGCGCACCCCGTTATTAAGATTGAATCAACACAGCG	900
Db	1028	CGCAGCCGCAAGGTGGCCCGGACGCGCGCACCCCGTTATTAAGATTGAATCAACACAGCG	1087
Oy	901	TTGACGCCCATTCACACCGCAAAAACAGCGAGTGGTGTGATTAATTAACCACTTCAGTGTG	960
Db	1088	TTGACGCCCATTCACACCGCAAAAACAGCGAGTGGTGTGATTAATTAACCACTTCAGTGTG	1147
Oy	961	TTTATGCGCGGACACGATTAATCTATTTGGCAATCTCGCGCGCGCACTGGAGCTCAACTG	1020
Db	1148	TTTATGCGCGGACACGATTAATCTATTTGGCAATCTCGCGCGCGCACTGGAGCTCAACTG	1207
Oy	1021	ACGCTTCCCGGACAGCGGATTAACACGCGCGCAGGTGGTGAATCTGTGTTGAACGCTGG	1080
Db	1208	ACGCTTCCCGGACAGCGGATTAACACGCGCGCAGGTGGTGAATCTGTGTTGAACGCTGG	1267
Oy	1081	CGTGGCGATTAAGGATTAACAGCCAGTGAATTCAGTTTGGCTGTGTCTTCAACTTTTACG	1140
Db	1268	CGTGGCGATTAAGGATTAACAGCCAGTGAATTCAGTTTGGCTGTGTCTTCAACTTTTACG	1327
Oy	1141	CAGATGCGGATTAACAGCGCGGTCAATTAATTAAGCGCGCCGGAGAGGTAAACTGAC	1200
Db	1328	CAGATGCGGATTAACAGCGCGGTCAATTAATTAAGCGCGCCGGAGAGGTAAACTGAC	1387
Oy	1201	CTGGCAGGATGTGAAGAGCGGAATGCGACAGGAGCTGTGTTGTTGGCAGGTTTACGCA	1266
Db	1388	CTGGCAGGATGTGAAGAGCGGAATGCGACAGGAGCTGTGTTGTTGGCAGGTTTACGCA	1447
Oy	1261	ATCTGGAATGAAGCAACGCTAACCCGCTTGCAATTTGTAA	1299
Db	1448	ATCTGGAATGAAGCAACGCTAACCCGCTTGCAATTTGTAA	1486

```

RESULT 14
ADO50295
ID      ADO50295 standard; DNA; 1901 BP
XX
AC      ADO50295;

```

Query	1	ARGAAGGCAATCTTATCCATTTTATCTCTTGATTCGGTAAACCCGCAATCTGCA	60
XX	29-JUN-2004	(first entry)	
XX	DE	Escherichia coli K12 appa phytase mutant DNA #1.	
XX	XX	Appa phytase; bacteria; thermal tolerance; protease stability; foodstuff;	
XX	KW	phytase; animal feed; fish feed; dough; baking; gene; mutant; ds.	
XX	OS	Escherichia coli; K12.	
XX	OS	Synthetic.	
XX	FH	Key	Location/Qualifiers
XX	FT	mutation	replace(389,T)
XX	FT	mutation	/*tag= a
XX	FT	mutation	replace(390,G)
XX	FT	mutation	/*tag= b
XX	PN	US2004091968-A1.	
XX	PD	13-MAY-2004.	
XX	PF	20-JUN-2003; 2003US-00601319.	
XX	PR	13-AUG-1997; 97US-00910798.	
XX	PR	01-MAR-1999; 99US-00259214.	
XX	PR	13-APR-1999; 99US-00291931.	
XX	PR	25-MAY-1999; 99US-00318528.	
XX	PR	24-MAY-2001; 2001US-00580515.	
XX	XX	(SHOR) SHORT J M.	
XX	PA	(KRET) KRETZ K.	
XX	PA	(GRAY) GRAY K A.	
XX	PA	(BART) BARTON N R.	
XX	PA	(GAR) GARRETT J B.	
XX	PA	(ODON) O'DONOGHUE E.	
XX	PA	(MATH) MATHER E J.	
XX	PI	Short JM, Kretz K, Gray KA, Barton NR, Garrett JB, O'donoghue E;	
XX	PI	Mather EJ;	
XX	DR	WPI: 2004-374952/35.	
XX	PT	Producing phytase, involves providing nucleic acid derived from bacteria	
XX	PT	encoding polypeptide having phytase activity, and expressing nucleic acid	
XX	PT	in yeast.	
XX	PS	Disclosure; SEQ ID NO 5; 74pp; English.	
XX	XX	The invention relates to a method for producing phytase that involves	
XX	CC	providing a nucleic acid encoding phytase derived from a bacteria, and	
XX	CC	expressing the nucleic acid in a yeast under conditions that allow	
XX	CC	expression of the enzyme in the yeast. The invention also relates to	
XX	CC	modified phytase enzyme which has improved thermal tolerance and protease	
XX	CC	stability at low pH. The phytase enzyme can be used in foodstuffs to	
XX	CC	improve the feeding value of phytase rich ingredients, and in diet of	
XX	CC	numerous animals including mammals, fowls and fishes, commercially	
XX	CC	significant mammals such as pigs, goats, laboratory rodents, commercially	
XX	CC	significant avian species such as chicken, ducks, doves, parrot, etc.,	
XX	CC	commercially farmed fish such as guppy, zebrafish, molly, swordtail,	
XX	CC	etc., in dough making and baking, in dietary aids for animals. The method	
XX	CC	provides easy manufacture of the active ingredient loaded biocompatible	
XX	CC	composition, higher yields and loading efficiency. The phytase	
XX	CC	incorporated in to the dietary aid is safe for animals. The present	
XX	CC	sequence is Escherichia coli K12 appa phytase mutant DNA.	
XX	SO	Sequence 1901 BP; 475 A; 499 C; 499 G; 427 T; 0 U; 1 Other;	
XX	Query Match	99.3%; Score 1290; DB 12; Length 1901;	
XX	Best Local Similarity	99.5%; Pred. No. 0;	
XX	Matches 1293; Conservative	0; Mismatches 6; Indels 0; Gaps 0	

```

Db      188 ATGAAAGGATCTTATCCATTTTATCTCTTCGATTCCTGTTAACCCCGCAATCTCA 247
Qy      61 TTGCTGAGTGAAGCCGAGAGCTGAAGCTGAAAGTGTGTGATTTGCACTGTCATGCT 120
Db      248 TTGCTGAGTGAAGCCGAGAGCTGAAGCTGAAAGTGTGTGATTTGCACTGTCATGCT 307
Qy      121 GTGCTGTCTCCAAACCAAGGCCCACTGATGACAGATGTCACCCGAGCATGAGCA 180
Db      308 GTGCTGTCTCCAAACCAAGGCCCACTGATGACAGATGTCACCCGAGCATGAGCA 367
Qy      181 ACCTGGCCGTTAAACTGGGTGGCTGACACCGCGCGGTGTGAGCTAATGCTATCTC 240
Db      368 ACCTGGCCGTTAAACTGGGTGAGCTGACACCGCGCGGTGTGAGCTAATGCTATCTC 427
Qy      241 GGACATTAACCAAGCCAGGCTGTGATGACCGAGATTCCTGGCGAATAAGGGCTGCCG 300
Db      428 GGACATTAACCAAGCCAGGCTGTGATGACCGAGATTCCTGGCGAATAAGGGCTGCCG 487
Qy      301 CAGTCTGTGAGTGGCGGATTTATGCTGATGTCGACGAGCGTAAACCGTAAACAGCGCA 360
Db      488 CAGTCTGTGAGTGGCGGATTTATGCTGATGTCGACGAGCGTAAACCGTAAACAGCGCA 547
Qy      361 GCCTTCGCGCGGGCTGACCTGACCTGCACTGTGCAATTAACCGTAAACAGCGCAATCG 420
Db      548 GCCTTCGCGCGGGCTGACCTGACCTGCACTGTGCAATTAACCGTAAACAGCGCAATCG 607
Qy      421 TCAGTCCCGATCCGTTATTTAATCCTCTAAACCTGGCTTGGCAATCGATTAACGG 480
Db      608 TCAGTCCCGATCCGTTATTTAATCCTCTAAACCTGGCTTGGCAATCGATTAACGG 667
Qy      481 AACGTGACGAGCGGATTCCTGACGAGGAGAGAGGTCATTAATGCTGACTTTACCGGGCAT 540
Db      668 AACGTGACGAGCGGATTCCTGACGAGGAGAGAGGTCATTAATGCTGACTTTACCGGGCAT 727
Qy      541 CCGCAACCGCGCTTTGCGCAACTGAAACGGGTCTTAATTTTCCGCAATCAAACTGTGTC 600
Db      728 CCGCAACCGCGCTTTGCGCAACTGAAACGGGTCTTAATTTTCCGCAATCAAACTGTGTC 787
Qy      601 CTTAAACGTGAGAAACAGACGAAAGCTGTTCAATTAACGAGCATTAACATGGAATC 660
Db      788 CTTAAACGTGAGAAACAGACGAAAGCTGTTCAATTAACGAGCATTAACATGGAATC 847
Qy      661 AAGGTAGAGGCGCAAGATGCTATTAACGAGGCGGTAGACCTGCAATCAATGCTGAG 720
Db      848 AAGGTAGAGGCGCAAGATGCTATTAACGAGGCGGTAGACCTGCAATCAATGCTGAG 907
Qy      721 GAGATATTTCTCTGCAACAGCAAGGAAATGCCGAGCCGGGATGGGAAAGATCACC 780
Db      908 GAGATATTTCTCTGCAACAGCAAGGAAATGCCGAGCCGGGATGGGAAAGATCACC 967
Qy      781 GATTCAACCAAGTGAACACCTTGTCTAATGTTGCAATTAACGCAATTTTATTTGCTACA 840
Db      968 GATTCAACCAAGTGAACACCTTGTCTAATGTTGCAATTAACGCAATTTTATTTGCTACA 1027
Qy      841 CGCAGCGCAGAGGTGGCCGCGACCGCGCAACCCGTTATTAATTTGATGAACAGAG 900
Db      1028 CGCAGCGCAGAGGTGGCCGCGACCGCGCAACCCGTTATTAATTTGATGAACAGAG 1087
Qy      901 TTGACGCCCCATCCACCGCAAAAACAGCGGTATGTTGATCATTAACCACTTCAAGTCTG 960
Db      1088 TTGACGCCCCATCCACCGCAAAAACAGCGGTATGTTGATCATTAACCACTTCAAGTCTG 1147
Qy      961 TTTATGCGCGGACACGATTAATCTGGCAAAATCTGGCGCGGCACTGAGGCTCACTGG 1020
Db      1148 TTTATGCGCGGACACGATTAATCTGGCAAAATCTGGCGCGGCACTGAGGCTCACTGG 1207
Qy      1021 AGCGTTCCCGGTAGCGCGATTAACAGCGCGCGAGTGTGAATCTGATTTGAACCTGG 1080
Db      1208 AGCGTTCCCGGTAGCGCGATTAACAGCGCGCGAGTGTGAATCTGATTTGAACCTGG 1267
Qy      1081 CGTCGGTAAAGCATTAACAGCATGTGATTCAGTGTTCCTTCAGACTTTAAG 1140

```

```

Db      1268 CGTCGCTAAGCGATTAACAGCAAGCATGATTCAGGTTTCGCTGTTCCAGACTTTACAG 1327
Qy      1141 CAGATGGGTGATTAAGCGCGCTGTCAATTAATACCGCGCCCGGAGAGTGAATGAC 1200
Db      1328 CAGATGGGTGATTAAGCGCGCTGTCAATTAATACCGCGCCCGGAGAGTGAATGAC 1387
Qy      1201 CTGCGAGATGTAAGAGCGCAAAATGCGCAGGCAATGTTGTTGCGAGCTTTTACGCA 1260
Db      1388 CTGCGAGATGTAAGAGCGCAAAATGCGCAGGCAATGTTGTTGCGAGCTTTTACGCA 1447
Qy      1261 ATCGTAATGAAGCAGCATACCGCTTTCAGCTTTGTA 1299
Db      1448 ATCGTAATGAAGCAGCATACCGCTTTCAGCTTTGTA 1486

RESULT 15
AD050296
ID AD050296 standard; DNA; 1901 BP.
XX
XX AD050296;
AC
XX
XX 29-JUN-2004 (first entry)
DE
XX Escherichia coli K12 appA phytase mutant DNA #2.
XX AppA phytase; bacteria; thermal tolerance; protease stability; foodstuff;
KW phytase; animal feed; fish feed; dough; baking; gene; mutant; ds.
XX Escherichia coli; K12.
OS Synthetic.
XX
XX
FH Key Location/Qualifiers
FT mutation replace(437,C)
FT FT /*tag= a
FT mutation replace(438,A)
FT FT /*tag= b
FT mutation replace(439,A)
FT FT /*tag= c
XX
XX US2004091968-A1.
XX
XX 13-MAY-2004.
XX
XX 20-JUN-2003; 2003US-00601319.
XX
XX 13-AUG-1997; 97US-00910798.
XX 01-MAR-1999; 99US-00259214.
XX 13-APR-1999; 99US-00291931.
XX 25-MAY-1999; 99US-00318528.
XX 24-MAY-2000; 2000US-00580515.
XX 24-MAY-2001; 2001US-00866379.
XX
XX (SHOR/) SHORT J M.
XX (KRET/) KRETZ K.
XX (GRAY/) GRAY K A.
XX (BART/) BARTON N R.
XX (GARR/) GARRETT J B.
XX (ODON/) O'DONOGHUE E.
XX (MATH/) MATHER E J.
XX
XX Short JM, Kretz K, Gray KA, Barton NR, Garrett JB, O'donoghue E;
XX Mather EJ;
XX
XX MPI; 2004-374952/35.
XX
XX Producing phytase, involves providing nucleic acid derived from bacteria
XX encoding polypeptide having phytase activity, and expressing nucleic acid
XX in yeast.
XX
XX Disclosure; SEQ ID NO 6; 74pp; English.
XX
XX The invention relates to a method for producing phytase that involves
XX providing a nucleic acid encoding phytase derived from a bacteria, and

```

expressing the nucleic acid in a yeast under conditions that allow expression of the enzyme in the yeast. The invention also relates to modified phytase enzyme which has improved thermal tolerance and protease stability at low pH. The phytase enzyme can be used in feedstuffs to improve the feeding value of phytate rich ingredients, and in diet of numerous animals including mammals, fowls and fishes, commercially. Significant mammals such as pigs, goats, laboratory rodents, commercially significant avian species such as chicken, ducks, doves, parrot, etc., commercially farmed fish such as guppy, zebrafish, molly, swordtail, etc., in dough making and baking, in dietary aids for animals. The method provides easy manufacture of the active ingredient loaded biocompatible composition, higher yields and loading efficiency. The phytase incorporated in to the dietary aid is safe for animals. The present sequence is Escherichia coli K12 appa phytase mutant DNA.

Sequence 1901 BP; 472 A; 498 C; 501 G; 429 T; 0 U; 1 Other;

Query Match 99.2%; Score 1288.4; DB 12; Length 1901;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1292; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

QY 1 ATGAAGCCATCTTAATCCCATTTTATCTCTGATTCGTTAACCCCGCAATCGCA 60
DB 188 ATGAAGCCATCTTAATCCCATTTTATCTCTGATTCGTTAACCCCGCAATCGCA 247
QY 61 TTGCGTCAAGTAGAGCCGAGCTGAAAGTGTGATTTGTCATGTCATGTCATGTC 120
DB 248 TTGCGTCAAGTAGAGCCGAGCTGAAAGTGTGATTTGTCATGTCATGTCATGTC 307
QY 121 GTTCGCTGCTCAACCAAGGCGCAACGTATGCAAGTCCCAAGCGATGCGCA 180
DB 308 GTTCGCTGCTCAACCAAGGCGCAACGTATGCAAGTCCCAAGCGATGCGCA 367
QY 181 ACCTGGCCGGTAAACCTGAGTGGCTGACACCGCGCGTGTGATGATTCGCTTATC 240
DB 368 ACCTGGCCGGTAAACCTGAGTGGCTGACACCGCGCGTGTGATGATTCGCTTATC 427
QY 241 GACATTATACCAAGCCGAGCTGTGTGAGCCGAGCGATTTGCTGCGAAGAAAGGCTGCCG 300
DB 428 GACATTATACCAAGCCGAGCTGTGTGAGCCGAGCGATTTGCTGCGAAGAAAGGCTGCCG 487
QY 301 CAGTCTGCTGAGTGTGCGGATTAATGCTGATGTGAGAGGATCCGTTAAACAGGCGAA 360
DB 488 CAGTCTGCTGAGTGTGCGGATTAATGCTGATGTGAGAGGATCCGTTAAACAGGCGAA 547
QY 361 GCTTTCGCGCCGCGGCTGAGACCTGACTGTGCAATTAACCGTACATACCCAGGCGATACG 420
DB 548 GCTTTCGCGCCGCGGCTGAGACCTGACTGTGCAATTAACCGTACATACCCAGGCGATACG 607
QY 421 TCCAGTCCGATCCGTTATTTAATCTCTTAAACCTGCGCTTTGCCAATGGAATACCGG 480
DB 608 TCCAGTCCGATCCGTTATTTAATCTCTTAAACCTGCGCTTTGCCAATGGAATACCGG 667
QY 481 AACGTGACTACCGGATCTCTACAGAGGCGAGAGGTCATTTGCTGACTTTACCGGGCAT 540
DB 668 AACGTGACTACCGGATCTCTACAGAGGCGAGAGGTCATTTGCTGACTTTACCGGGCAT 727
QY 541 CGGCAAAACGCGGTTTCGCAACTGAAACGAGGTGCTTAATTTTCGCAATCAACTGTGC 600
DB 728 CGGCAAAACGCGGTTTCGCAACTGAAACGAGGTGCTTAATTTTCGCAATCAACTGTGC 787
QY 601 CTTAAACGTGAGAAACAGAGCAAGAACTGTTCAATTAACGAGGCAATTAACATCGAAATC 660
DB 788 CTTAAACGTGAGAAACAGAGCAAGAACTGTTCAATTAACGAGGCAATTAACATCGAAATC 847
QY 661 AAGGTGAGCGCGGCAATGTCATTAACCGGTGCGTAAGCTTCGATCAATGCTGACG 720
DB 848 AAGGTGAGCGCGGCAATGTCATTAACCGGTGCGTAAGCTTCGATCAATGCTGACG 907
QY 721 GAGATATTTCTCTGCAACAGACACAGGGAATGCGGAGCGGGGTGGGGAAGATCAC 780
DB 908 GAGATATTTCTCTGCAACAGACACAGGGAATGCGGAGCGGGGTGGGGAAGATCAC 967

```

```

QY 781 GATTACACACAGAGAAACACCTTGTAAGTTTGCAATACCGCAATTTTATTTGCTACAA 840
DB 968 GATTACACACAGAGAAACACCTTGTAAGTTTGCAATACCGCAATTTTATTTGCTACAA 1027
QY 841 CGCACCCAGAGGTTGCGCGACGCGCGCACCCGTTATTTAGATTGATCAAGACAGCG 900
DB 1028 CGCACCCAGAGGTTGCGCGACGCGCGCACCCGTTATTTAGATTGATCAAGACAGCG 1087
QY 901 TTGACCCCGCATTCACCGCAAAACAGGGGTATGTTGATCAATTAACCTTCAGGCTG 960
DB 1088 TTGACCCCGCATTCACCGCAAAACAGGGGTATGTTGATCAATTAACCTTCAGGCTG 1147
QY 961 TTATGCGCGGACACGATACTAATCTGCAAAATCTGCGCGCGCATGAGCTCAACTG 1020
DB 1148 TTATGCGCGGACACGATACTAATCTGCAAAATCTGCGCGCGCATGAGCTCAACTG 1207
QY 1021 ACCTTCCCGGTACGCGGATTAACACCGCGCGAGTGTGAACTGTTGTAACGCTGG 1080
DB 1208 ACCTTCCCGGTACGCGGATTAACACCGCGCGAGTGTGAACTGTTGTAACGCTGG 1267
QY 1081 CGTGGCTAAGCATTAACAGCCAGTGTGATTCAGGTTTCCGTGCTTCCAGACTTTAC 1140
DB 1268 CGTGGCTAAGCATTAACAGCCAGTGTGATTCAGGTTTCCGTGCTTCCAGACTTTAC 1327
QY 1141 CAGATGCGATTAACAGCCGCTGTCTATTAAATACGCGCGCGAGAGTGAACCTGACC 1200
DB 1328 CAGATGCGATTAACAGCCGCTGTCTATTAAATACGCGCGCGAGAGTGAACCTGACC 1387
QY 1201 CTGGCAGATGTGAAGAGCGAAATGCGCAGAGGATGTTGTTGCGAGGTTTTCGCAA 1260
DB 1388 CTGGCAGATGTGAAGAGCGAAATGCGCAGAGGATGTTGTTGCGAGGTTTTCGCAA 1447
QY 1261 ATGCTGATGAAGACAGCATACCCGCTTGCAAGTTTGTAA 1299
DB 1448 ATGCTGATGAAGACAGCATACCCGCTTGCAAGTTTGTAA 1486

```

Search completed: February 10, 2005, 09:41:14
Job time : 889.787 secs

THIS PAGE BLANK (USPTO)

QY 241 AGGTTGTTCTGTGGGACAGTAGAAATCGAAGGCGATGTGCCCTTTAGTGAACCTGATGG 300
DB 241 AGGGTGGTTCTGTGGGACAGTAGAAATCGAAGGCGATGTGCCCTTTAGTGAACCTGATGG 300
QY 301 AAGCTACCCCAACAGACAGATGTGTCACTCAGGCGCAATCCGTTTCGAGTTTGATGGCA 360
DB 301 AAGCTACCCCAACAGACAGATGTGTCACTCAGGCGCAATCCGTTTCGAGTTTGATGGCA 360
QY 361 GCCGGACAGTGAAGACAGACACCTTACTGATTGAGAGAGATAGAACTATGCGCAGC 420
DB 361 GCCGGACAGTGAAGACAGACACCTTACTGATTGAGAGAGATAGAACTATGCGCAGC 420
QY 421 AGGATTGAGAGACCTTGAACAGGCGGCAAGCCCTTAACACACACCTTACCTCACTTGG 480
DB 421 AGGATTGAGAGACCTTGAACAGGCGGCAAGCCCTTAACACACACCTTACCTCACTTGG 480
QY 481 ACAAAGCTGCCAAGACCA 500
DB 481 ACAAAGCTGCCAAGACCA 500

RESULT 2
AX042372 20623 bp DNA linear PAT 23-NOV-2000
LOCUS Sequence 1 from Patent WO0064247.
DEFINITION AX042372
ACCESSION AX042372
VERSION AX042372.1 GI:11340990
KEYWORDS
SOURCE
ORGANISM
synthetic construct
other sequences; artificial sequences.
REFERENCE
1 Forsberg, C.W., Golovan, S. and Phillips, J.P.
TITLE Transgenic animals expressing salivary proteins
JOURNAL Patent: WO 0064247-A 1 02-NOV-2000;
UNIVERSITY OF GUELPH (CA)

FEATURES
source
1..20623
Location/Qualifiers
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Tama2/APPA plasmid"

ORIGIN

Query Match 100.0%; Score 500; DB 6; Length 20623;
Best Local Similarity 100.0%; Pred. No. 1.1e-153;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TCGAGAGTATCTTTGTACAGTGTGCTCCACAAAGGGTACTGTTGCCACATAGAAG 60
DB 1 TCGAGAGTATCTTTGTACAGTGTGCTCCACAAAGGGTACTGTTGCCACATAGAAG 60
QY 61 ATCTAACTAATTAATTAATTCCTCACCAGCAATCTTTCAGTCACTAAGTTAGCAGAT 120
DB 61 ATCTAACTAATTAATTAATTCCTCACCAGCAATCTTTCAGTCACTAAGTTAGCAGAT 120
QY 121 TGTGTAACAAGTTCTCCAAAGAGAGATACAGATGATGCGGTATAGGGTGGACCTGGCTG 180
DB 121 TGTGTAACAAGTTCTCCAAAGAGAGATACAGATGATGCGGTATAGGGTGGACCTGGCTG 180
QY 181 CTGAGAGACACCTGATCTGACTAAGAGAGCAAGGTGTAAGTGAATGGTGGAGT 240
DB 181 CTGAGAGACACCTGATCTGACTAAGAGAGCAAGGTGTAAGTGAATGGTGGAGT 240
QY 241 AGGGTGTCTGTGGGACAGTAGAAATCGAAGGCGATGCGGTTTAACTGATGATGG 300
DB 241 AGGGTGTCTGTGGGACAGTAGAAATCGAAGGCGATGCGGTTTAACTGATGATGG 300
QY 301 AAGCTACCCCAACAGACAGATGTGTCACTCAGGCGCAATCCGTTTCGAGTTTGATGGCA 360
DB 301 AAGCTACCCCAACAGACAGATGTGTCACTCAGGCGCAATCCGTTTCGAGTTTGATGGCA 360

QY 361 GCCGACAGTGAAGACAGACACCTTACTGATTGAGAGAGATAGAACTATGCGCAGC 420
DB 361 GCCGACAGTGAAGACAGACACCTTACTGATTGAGAGAGATAGAACTATGCGCAGC 420
QY 421 AGGATTGAGAGACCTTGAACAGGCGGCAAGCCCTTAACACACACCTTACCTTGG 480
DB 421 AGGATTGAGAGACCTTGAACAGGCGGCAAGCCCTTAACACACACCTTACCTTGG 480
QY 481 ACAAAGCTGCCAAGACCA 500
DB 481 ACAAAGCTGCCAAGACCA 500

RESULT 3
AL732466 172535 bp DNA linear ROD 24-OCT-2002
LOCUS Mouse DNA sequence from clone RP23-149G14 on chromosome 2, complete
DEFINITION AL732466
ACCESSION AL732466
VERSION AL732466.7 GI:24394996
KEYWORDS
SOURCE
ORGANISM
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 172535)
REFERENCE
1 Sycamore, N.
TITLE Direct Submission
JOURNAL Submitted (20-OCT-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Oct 25, 2002 this sequence version replaced gi:21621699.
COMMENT
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WormPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-149G14 is from the RP23-23 Mouse PAC library constructed by the group of Pieter de Jong. For further details see <http://www.choi.org/bacpac/home.htm> VECTOR: pBACe3.6.

FEATURES

source
1..172535
Location/Qualifiers
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="2"
/clone="RP23-149G14"
/clone_lib="RP23-23"

ORIGIN

Query Match 100.0%; Score 500; DB 10; Length 172535;
Best Local Similarity 100.0%; Pred. No. 1.6e-153;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 TCGAGAGTATCTTTGTACCTGTGCTCCAAAGGGGTACTGTTGCCCATAGAAAG 60
Db 19120 TCGAGAGTATCTTTGTACCTGTGCTCCAAAGGGGTACTGTTGCCCATAGAAAG 19179
QY 61 ATCTAAATCTTATTAATTAATCCCTCCACCGGCAATCTTTCAGTCACTAAGTAGACGAT 120
Db 19180 ATCTAAATCTTATTAATTAATCCCTCCACCGGCAATCTTTCAGTCACTAAGTAGACGAT 19239
QY 121 TGTGTAACAGATCTTCCAAAGAGAGATACAGATGATGCGGTATAGAGGTGACCTGCTG 180
Db 19240 TGTGTAACAGATCTTCCAAAGAGAGATACAGATGATGCGGTATAGAGGTGACCTGCTG 19299
QY 181 CTGAGAGACACCTGATCTGACTTAAGAAAGGACCGGTGTATTGATGATGTGTGAGT 240
Db 19300 CTGAGAGACACCTGATCTGACTTAAGAAAGGACCGGTGTATTGATGATGTGTGAGT 19359
QY 241 AGGTGTGTTCTGTGGGACAGTACGATAAATCGAGAGGATGTCGGTTTATGAACTGATG 300
Db 19360 AGGTGTGTTCTGTGGGACAGTACGATAAATCGAGAGGATGTCGGTTTATGAACTGATG 19419
QY 301 AAGCTACCCCAAGACAGACAGATGTCAGTCAAGGCCATCCGTTGAGTTGATGGGCA 360
Db 19420 AAGCTACCCCAAGACAGACAGATGTCAGTCAAGGCCATCCGTTGAGTTGATGGGCA 19479
QY 361 GCCGACAGTACAGACACACCTACTACTGTTGAGAGAGAGTAGAGAACATAGGCCAGC 420
Db 19480 GCCGACAGTACAGACACACCTACTACTGTTGAGAGAGAGTAGAGAACATAGGCCAGC 19539
QY 421 AGGATTTGAGAGACCTTGAAGGGCGAAGGCCCTTAACACACACCTACCTCACTTG 480
Db 19540 AGGATTTGAGAGACCTTGAAGGGCGAAGGCCCTTAACACACACCTACCTCACTTG 19599
QY 481 ACAAGCTGCGCAAGACCA 500
Db 19600 ACAAGCTGCGCAAGACCA 19619

RESULT 4
AC107726 214345 bp DNA linear HTG 03-MAR-2003
LOCUS Mus musculus clone RP24-85E12, WORKING DRAFT SEQUENCE, 8 unordered
DEFINITION pieces.
AC107726 GI:2861318
VERSION HTG; HTGS PHASE1; HTGS DRAFT.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 214345)
2 (bases 1 to 214345)
Mus musculus, clone RP24-85E12
Unpublished
2 (bases 1 to 214345)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhalter, B.,
Brown, A., Camarata, J., Campolano, A., Chang, J., Chazaro, B.,
Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Larocque, K., Lamazares, R.,
Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C.,
Macdonald, P., Major, J., Marguis, N., Matthews, C., McCarthy, M.,
McWan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T.,
Menga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,
Reita, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,
Rosetti, M., Roy, A., Santos, R., Schauer, S., Schnupack, R., Seaman, S.,
Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Straus, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 214345)
Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,
Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T.,
Boguslavsky, L., Boukhalter, B., Camarata, J., Chang, J., Choepel, Y.,
Collymore, A., Cook, A., Cooke, P., Corum, B., DeArrellano, K.,
Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,
Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S.,
Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,
Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,
Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C.,
Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,
Meldrum, J., Meneus, L., Mihova, T., Menga, V., Murphy, T., Naylor, J.,
Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
Rachupka, A., Ramasamy, U., Raymond, C., Reita, R., Rise, C., Rogov, P.,
Roman, J., Schauer, S., Schnupack, R., Seaman, S., Severy, P., Smith, C.,
Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,
Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,
Wassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (03-MAR-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 3, 2003 this sequence version replaced gi:28173217.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L20103
Center clone name: 85 E 12
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 21253 bases at least Q40
Consensus quality: 21324 bases at least Q30
Consensus quality: 21351 bases at least Q20
Insert size: 19400; agarose-fp
Insert size: 213645; sum-of-coverage
Quality coverage: 13.0 in Q20 bases; agarose-fp
Quality coverage: 11.8 in Q20 bases; sum-of-coverage
-----
NOTE: This is a 'working draft' sequence. It currently
consists of 8 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
1
64960: contig of 64960 bp in length
64961: gap of 100 bp
65061: contig of 3215 bp in length
68275: gap of 100 bp
68276: gap of 100 bp
68376: contig of 6133 bp in length
74568: gap of 100 bp
74569: gap of 100 bp
74668: contig of 6353 bp in length
81021: gap of 100 bp
81022: gap of 100 bp
81121: contig of 10329 bp in length
81132: gap of 100 bp
91450: contig of 10329 bp in length
91451: gap of 100 bp

```

* 91551 118992: contig of 27442 bp in length
* 118993 119092: gap of 100 bp
* 119093 165215: contig of 46123 bp in length
* 165215 165315: gap of 100 bp
* 165316 214345: contig of 49030 bp in length.

FEATURES

SOURCE

1. 214345
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP24-85E12"
/clone_lib="RP24-85E12 Male Mouse BAC"
1. 64960
/note="assembly_fragment"
vector side:left"
clone_end:SP6
misc_feature
65061..68275
/note="assembly_fragment"
misc_feature
68376..74568
/note="assembly_fragment"
misc_feature
74669..81021
/note="assembly_fragment"
misc_feature
81122..91450
/note="assembly_fragment"
misc_feature
91551..118992
/note="assembly_fragment"
misc_feature
119093..165215
/note="assembly_fragment"
misc_feature
165316..214345
/note="assembly_fragment"
clone_end:T7
vector_side:right"

ORIGIN

Query Match 100.0%; Score 500; DB 2; Length 214345;

Best Local Similarity 100.0%; Pred. No. 1,7e-153;
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 TCGAGAGTATCTTTGTGACGTGCTCCCAAGAGGGTACTGTTCCCACTAGAAAG 60
49912 TCGAGAGTATCTTTGTGACGTGCTCCCAAGAGGGTACTGTTCCCACTAGAAAG 49971
61 ATCTAATCAATTAATTAATCCCTCAACCGGCAATCTTTCAGTCATTAAGTTCAGCAT 120
49972 ATCTAATCAATTAATTAATCCCTCAACCGGCAATCTTTCAGTCATTAAGTTCAGCAT 50031
121 TGTGAACAAGTCTCTCAAGAGAGATACAGATGAGTGCATAGAGGTGACCTGCTG 180
50032 TGTGAACAAGTCTCTCAAGAGAGATACAGATGAGTGCATAGAGGTGACCTGCTG 50091
181 CTGAGAGACACTGTCATCTGACTTAAGAGAGCCGCGTGTAGTTGATGCTGTGAGT 240
50092 CTGAGAGACACTGTCATCTGACTTAAGAGAGCCGCGTGTAGTTGATGCTGTGAGT 50151
241 AGGATGTTCTGAGGACAGTAGAATAAGAGAGGATGAGCGTTTAACTAGTAATGATG 300
50152 AGGATGTTCTGAGGACAGTAGAATAAGAGAGGATGAGCGTTTAACTAGTAATGATG 50211
301 AAGCTACCCCAACGACAGAGATGTGACGTCAAGGCCAATCGTTTCGAGTTTGAAGGGCA 360
50212 AAGCTACCCCAACGACAGAGATGTGACGTCAAGGCCAATCGTTTCGAGTTTGAAGGGCA 50271
361 GCGGACAGTGAACAGACACACTTACTGATGTTGAGGAGAGTAGAAGAAAGAGGCGCAGC 420
50272 GCGGACAGTGAACAGACACACTTACTGATGTTGAGGAGAGTAGAAGAAAGAGGCGCAGC 50331
421 AGGATGAGAGACCTGACAGGCGCAAGGCCCTTAACACACACACTTACCACTCACTTG 480
50332 AGGATGAGAGACCTGACAGGCGCAAGGCCCTTAACACACACACTTACCACTCACTTG 50391
481 ACAAGCTGCCAAAGACCAA 500
50392 ACAAGCTGCCAAAGACCAA 50411

RESULT 5
AC103277/c
LOCUS
DEFINITION
AC103277 236802 bp DNA linear HTG 13-MAY-2003
Rattus norvegicus clone CH230-235G1, *** SEQUENCING IN PROGRESS
***, 2 unordered pieces.

AC103277.6 GI:30579004
HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
Rattus norvegicus (Norway rat)
SOURCE
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE

AUTHORS

1 (bases 1 to 236802)
Munzy,D,Marie, Metzker,M, Lee, Abramson,S, Adams,C, Alder,J, Allen,C, Allen,H, Alebrooks,S, Amin,A, Angiano,D, Anyalebechi,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H, Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benahmed,F, Biewald,K, Blair,J, Blankenburg,K, Blyth,P, Brown,M, Bryant,N, Buhay,C, Burch,P, Burrell,K, Calderon,E, Cardenas,V, Carter,K, Cavazos,I, Cesaar,H, Center,A, Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J, Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L, Davila,M, Davis,C, Davy-Carroll,L, De Anda,C, Dederich,D, Delgado,O, Denson,S, Deramo,C, Ding,Y, Dinh,H, Divya,K, Draper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Evans,K, Egan,A, Escotto,M, Eugene,C, Evans,C, Falls,T, Fan,G, Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P, Frazer,C,M, Gabisi,A, Ganta,R, Garcia,A, Garner,T, Garza,W, Gebregeorgis,E, Geer,K, Gill,R, Garcia,M, Guerra,W, Guevara,M, Gunaratne,P, Haaland,W, Hamil,C, Hamilton,C, Hamilton,K, Harvey,Y, Haylak,P, Hawes,A, Henderson,N, Hernandez,J, Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hogues,M, Hollins,B, Howells,S, Huliy,S, Hume,J, Idlebird,D, Jackson,A, Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jolivet,A, Karpathy,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovar,C, Kowis,C, Kraft,C,L, Lebow,H, Levan,J, Lewis,L, Li,Z, Liu,J, Liu,J, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,J, Lorenshewa,L, Louissege,H, Lozano,R,D, Lu,X, Ma,J, Maheshwari,M, Mahindratne,M, Mahmood,M, Malloy,K, Mangum,A, Mangum,B, Mapua,P, Martin,K, Martin,R, Martinez,E, Mahoney,S, McLeod,M,P, McNeill,T,Z, Meenen,E, Milosavljevic,A, Miner,G, Minja,E, Montemayor,J, Murphy,M, Morgan,M, Morris,K, Morris,S, Mundana,M, Murphy,M, Natr,L, Nankervis,C, Neal,D, Newton,N, Nguyen,N, Norris,S, Nwackekeh,O, Okwionu,G, Olariupana,A, Pal,S, Parks,K, Pasternak,S, Paul,H, Perez,A, Perez,L, Pfannkuch,C, Plapper,F, Poindexter,A, Popovic,D, Primus,B, Pu,L, Puzo,M, Qutroz,J, Rachlin,E, Reeves,K, Regier,M,A, Reigh,R, Reilly,B, Reilly,M, Ren,Y, Reuter,M, Richards,S, Riggs,F, Rivers,C, Rodkey,T, Rojas,A, Rose,M, Rose,R, Ruiz,S, Sanders,M, Savery,G, Scherer,S, Scott,G, Shatman,S, Shen,H, Shetty,J, Shvartsbeyn,A, Sleson,I, Sletter,C,D, Smajic,D, Sneed,A, Sodergren,E, Song,X-Z, Sorelle,R, Sosa,J, Steimle,M, Strong,R, Sutton,A, Svattek,A, Tabors,P, Taylor,C, Taylor,T, Thomas,N, Thomas,S, Tingey,A, Trejos,Z, Uemari,K, Valas,R, Vere,V, Villasana,D, Waldron,L, Walker,B, Wang,J, Wang,Q, Wang,S, Warren,J, Warren,R, Wei,X, White,F, Williams,G, Willson,R, Wleczky,R, Wooden,H, Wotley,K, Wright,D, Wright,R, Wu,J, Yakub,S, Yen,J, Yoon,L, Yoon,V, Yu,F, Zhang,J, Zhou,J, Zhou,X, Zhao,S, Dunn,D, von Niederhausern,A, Weiss,R, Smith,D,R, Holt,R,A, Smith,H,O, Weinstock,G, and Gibbs,R.A.

TITLE

Unpublished

Unpublished

Unpublished

Unpublished

Unpublished

Unpublished

Unpublished

Submitted (24-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

Matches	346;	Conservative	0;	Mismatches	83;	Indels	40;	Gaps	7;
QY	3	GAGAGTATCTTTGTCAGCTGTGCTCTCCAAAGAGGGGATCTGTCCAC--ATGAAGA	60						
Db	180303	GGGAGGATCTTTGTCAGCTGTGCAATCCAAAGAGATGTTGTAACCATGAGAGAGAG	180244						
QY	61	ATCTAACTATTAATTAATCCCTCACCCGGAATCTTTCAGTCACTAATGAGACGAT	120						
Db	180243	AGCTAAACAAATTAATTA--CCCCCCCCCAATCTTTCAGTCACTAATGAGAGAT	180185						
QY	121	TGTTGAACAAGTCTCCAAAGAGAGATACAGATAGTGCCTATAG--GGTGACC----	174						
Db	180184	TGTTGAATTAATTTCTCCAAAGAGAAATGCAAGAGGGGCTCTCTGTAGTGGACCCGGC	180125						
QY	175	-----TGCTGTGAGAGAGACACTCGATCTGACTAGAGACCAAGGTATTAGT	225						
Db	180124	ACCGTTTATGCTGTCTTAAGAGACCTCTGCACTGACTAAGAGAGGCGACATATTTAGT	180065						
QY	226	TGAATGCTGTGAGAGAGAGGTGTGTTCTGTGGGACATAGAAATGAGAGGCATGTGCGCT	285						
Db	180064	CAAA-----TGTTCTGTGGAGAGATTAATAAATGAGAGGACAGTCCAT	180020						
QY	286	TTAGTGAAGTATGAGAGACTACCCCAAGACAGAGATTGTGATCAGGCGCAATCCGTTT	345						
Db	180019	TTAGCAAACTATGAGAACTATCTCGAATGTTCAGGAGTATCAATAGGCGCATGTAATTT	179966						
QY	346	CGAGTTGTATGGCGAGCCGCGACATGAGACAGACACACTCACTAGTGG--CGAAGA	403						
Db	179959	CAAGTTATGGCGAGCTCAGATATGTGGGACAGATCACCTGCACAGTTGGAAGGGAGGA	179900						
QY	404	TGAGAA-----CAATGCCAGACAGGAGATTGAGAGACCCGGAAGGGCA	447						
Db	179899	GAAAGATGAGGGAGAGGGCGACAGCAACGACCAAGAGACCTTAACAGGTGCA	179851						
RESULT 6									
CRUSMT3A	LOCUS	CRUSMT3A	915 bp	mRNA	linear	ROD 01-DEC-2003			
DEFINITION		Cricetulus griseus MIF2 suppressor (SMT3) mRNA, complete cds.							
ACCESSION		L79950							
VERSION		L79950.1 GI:38570349							
KEYWORDS		MIF2 suppressor.							
SOURCE		Cricetulus griseus (Chinese hamster)							
ORGANISM		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.							
REFERENCE		1 (bases 1 to 915)							
AUTHORS		Mannem,H., Tsol,S.C., Krushkal,J.S., Li,W.H. and Li,S.S.							
TITLE		The cDNA cloning and molecular evolution of reptile and pigeon lactate dehydrogenase isozymes							
JOURNAL		Mol. Biol. Evol. 14 (11), 1081-1087 (1997)							
MEDLINE		98031373							
PUBMED		9364765							
FEATURES		source							
		1..915							
		/organism="Cricetulus griseus"							
		/mol_type="mRNA"							
		/db_xref="taxon:10029"							
		/cell_line="CHO-K1 G0 phase"							
		/tissue_type="ovary"							
		1..915							
		/gene="SMT3"							
		1..33							
		/gene="SMT3"							
		/note="putative"							
		34..321							
		/gene="SMT3"							
		/note="putative"							
		/codon_start=1							
		/product="MIF2 suppressor"							
		/protein_id="AAR24618.1"							
		/db_xref="GI:38570350"							
gene									

```

/translation="MADEKPKGKVTENNNDHNLKVAGDGSVVOFKIKRHPLSKLM
KAYCERGLSMKQIFRFDGQINETDPAOLENDEDEDTIVFOQOTGVV"
3'UTR
/gene="SMT3"
/size="915"
/note="putative"
polyA_signal
/gene="SMT3"
/size="900..905"
/note="putative"
polyA_site
/size="915"
/note="SMT3"
/note="putative"
ORIGIN
Query Match      19.3%; Score 96.6; DB 10; Length 915;
Best Local Similarity 63.7%; Pred. No. 2.8e-20;
Matches 179; Conservative 0; Mismatches 99; Indels 3; Gaps 2;

QY 221 TTAGTTGATGGTGTGAGTGAAGGTTCTGTGTGGGACAGTAAATCGAGAGCATGT 280
DB 87 TAAATTGAAGGTGGCGGGGAGAGATGTTCTGTGTGGGAGTTAAGATTAAAGGCATAC 146
QY 281 GCCGTTTGAAGTGAAGTGAAGG-CTACCCCAAGCAGAGATTGTCTAGTGGCCAT 339
DB 147 ACCACTTAGTAACATTAAGAAAGCCATTTGTGAACGACAGAGATTGTCAATGAGCGAGAT 206
QY 340 CCGTTTCGAGTTTGAAGTGGGACCGGACAGTGAACAGACACCTACTCAGTTGGAGGA 399
DB 207 CAGATTCCCGTTTGAAGTGAAGCAGCAATCAATGAACAGACCTGCAAGTTGGAAT 266
QY 400 AGATGAGAACAATGGCCAGCAGGAGTGAAGACCCCTGACAGGCGCAAGCCCTAACAC 459
DB 267 GGAGATGAAGATACATTTGATGTG--TTCCAGACAGACAGAGAGGTCTACTTAATAA 324
QY 460 ACACACCTACCACTCTTGTGACAAAGTGCACCAAGACCA 500
DB 325 GGGACCTGCTACTTACTTCAGAAATTTGTTATAGACCA 365

RESULT 7
AC119067 193199 bp DNA linear HTG 14-JAN-2003
LOCUS Canis familiaris clone Rp81-23M24, WORKING DRAFT SEQUENCE, 11
DEFINITION
ORDERED PIECES.
AC119067
AC119067.4 GI:27733945
HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE
Canis familiaris
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
1 (bases 1 to 193199)
Ahter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
Benjamin,B., Blakesley,R.W., Boufield,G.G., Brinkley,C., Brooks,S.,
Carliga,K., Coleman,B., Engle,J., Granite,S., Guan,X., Gupta,J.,
Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E.,
Latic,P., Lee-Lin,S.-O., Legaspi,R., Maduro,O.L., McDowell,J.,
Margulies,E.H., Masello,C., Maerki,B., McDowell,J.,
Regulirgen,C., Pearson,R., Portnoy,M.E., Prasad,A.,
Reddix-Dugue,N., Schandler,K., Schuler,M.G., Simon,C.,
Stanthorp,S., Thomas,J.W., Thomas,P.J., Touchman,J.W., Vogt,J.L.,
Wetherby,K.D., Wiggins,L., Young,A. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 193199)
Green,E.D.
Direct Submission
Submitted (24-APR-2002) NIH Intramural Sequencing Center, 8717
Groveomont Circle, Gaithersburg, MD 20877, USA
3 (bases 1 to 193199)
Green,E.D.
Direct Submission
Submitted (14-JAN-2003) NIH Intramural Sequencing Center, 8717
Groveomont Circle, Gaithersburg, MD 20877, USA

```

COMMENT

On Jan 14, 2003 this sequence version replaced gi:24308653.
 ----- Genome Center
 Center: NIH Intramural Sequencing Center
 Center code: NISC
 Web site: <http://www.nisc.nih.gov>
 Contact: nisc_zoo@nhgri.nih.gov
 ----- Project Information
 Center project name: dga
 Center clone name: 023M24

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8x average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 19149 bases at least Q40
 Consensus quality: 191542 bases at least Q30
 Consensus quality: 191909 bases at least Q20
 Insert size: 188000; agarose-fp
 Insert size: 192199; sum-of-contigs
 Quality coverage: 10.74x in Q20 bases; agarose-fp
 Quality coverage: 10.50x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 11 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

* This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.

```

1 20054: contig of 20054 bp in length
* 20055 20154: gap of unknown length
* 20155 23880: contig of 3726 bp in length
* 23881 23980: gap of unknown length
* 23981 29813: contig of 5833 bp in length
* 29814 29913: gap of unknown length
* 29914 60037: contig of 30124 bp in length
* 60038 72455: gap of unknown length
* 72455 72555: gap of unknown length
* 72556 74766: contig of 2211 bp in length
* 74767 74866: gap of unknown length
* 74867 126334: contig of 51468 bp in length
* 126335 126434: gap of unknown length
* 126435 140029: contig of 13595 bp in length
* 140030 140129: gap of unknown length
* 140130 144562: contig of 4433 bp in length
* 144563 144662: gap of unknown length
* 144663 183707: contig of 39045 bp in length
* 183708 183807: gap of unknown length
* 183808 193199: contig of 9392 bp in length.

```

FEATURES

source

```

1..193199
/organism="Canis familiaris"
/mol_type="genomic DNA"
/db_xref="taxon:9615"
/clone_lib="Rp81-23M24"
1..20054
/note="assembly_fragment"
misc_feature

```

clone_end:Sp6
vector_side:left"
misc_feature
2015..23880
/note="assembly_fragment"
misc_feature
23981..29813
/note="assembly_fragment"
misc_feature
29914..60037
/note="assembly_fragment"
misc_feature
60138..72455
/note="assembly_fragment"
misc_feature
72556..74766
/note="assembly_fragment"
misc_feature
74867..126334
/note="assembly_fragment"
misc_feature
125898..193199
/note="clone overlaps with GenBank Accession Number
AC120521 clone RP81-392C9 (center project name dgb)"
misc_feature
126435..140029
/note="assembly_fragment"
misc_feature
140130..144562
/note="assembly_fragment"
misc_feature
144663..183707
/note="assembly_fragment"
misc_feature
193808..193199
/note="assembly_fragment"
clone_end:T7
vector_side:right"

ORIGIN

Query Match 18.5%; Score 92.4; DB 2; Length 193199;
Best Local Similarity 65.9%; Pred. No. 1.7e-18;
Matches 149; Conservative 0; Mismatches 76; Indels 1; Gaps 1;
QY 206 AGAAGACCCAGGTGTAGTGAATGCTGTGAGTGGTGTCTGTGGACAGTAGAA 265
58554 AGAAGACCCAGTATTTATCTGAAGTGGGGGCGAGATGTTCTATGTGACAGTTAA 58613
QY 266 AATCGAGAGGCGATGTCGTTAGTGAATGATGATGAG-CTACCCCAAACAGACAGATT 324
58614 GATTAGAGGCGATTCACATTTAGTAACTAATGAAAGCCTACTGTGTAACGACAGGCTTT 58673
QY 325 GTCAAGTACGCCATCCGTTTCCAGTTTGAATGGGCGACCCGACAGTGACAGACACACC 384
58674 GTCAATGAGGCGATTCGATTCGGGGTGAATGGGCAACCAATCAATAAACAGACACACC 58733
QY 385 TACTCAGTTGAGAGATGAGAACATGCGCCAGCAGGATTGAG 430
58734 TGCAAGCTGGAATGAGAAAGAGACAACTGCTGTTCAG 58779
DB

RESULT 8
PTU89438 717 bp mRNA linear VRL 10-JAN-2000
LOCUS PTU89438
DEFINITION Pestivirus type 1 nonstructural protein p125-2 (p125) mRNA, partial
cDNA.
ACCESSION U89438
VERSION U89438.1 GI:2707596
KEYWORDS Bovine viral diarrhea virus 1 (BVDV-1)
SOURCE Bovine viral diarrhea virus 1
ORGANISM Pestivirus; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Pestivirus.
REFERENCE 1 (bases 1 to 717)
AUTHORS Qi, F., Ridpath, J.F. and Berry, E.S.
TITLE Insertion of a bovine SM7B gene in NS4B and duplication of NS3 in
a bovine viral diarrhea virus genome correlate with the
cytopathogenicity of the virus
JOURNAL MEDLINE
MEDLINE 99049626
PUBMED 9833880
REFERENCE 2 (bases 1 to 717)
AUTHORS Qi, F. and Berry, E.S.
TITLE Direct Submission

JOURNAL Submitted (12-FEB-1997) Oral Biology, University of Alabama at
Birmingham, 250 LHRB 1919 7th Ave. S, Birmingham, AL 35294, USA
FEATURES
source
location/Qualifiers
1..717
/organism="Bovine viral diarrhea virus 1"
/mol_type="mRNA"
/strain="SNC"
/db_xref="taxon:11099"
/note="p125 region"
1..717
/gene="p125"
1..717
/gene="p125"
CDS
1..717
/gene="p125"
/note="the 3' region of the p125 protein is duplicated in
this strain. Therefore, U89437 is the first copy, and the
3' region of the U89438 sequence contains the second copy"
/codon_start=1
/product="nonstructural protein p125-2"
/protein_id="AAB92355.1"
/db_xref="GI:2707597"
/translation="PWITDIYITIDQRLDPTHTLOYAPNAIRTEGKETELKELAVGL
DKIMEYLEHLVMAADSKPEKGVKTENNHNINAKVAGQDSVVOFKIKHTPTLSKIMKA
YCEKQGLSMRQIRFRPDGQINETDTPALEDEDTIDVFOQTGGGAVCKKITEH
EKCHVNIIDKLTAFPGVMRGTPPAPVPFPAALIKVRGLEGTMAVTHQGISVDH
VTAKGLDIVCDSMGRTRVVCQ"

ORIGIN

Query Match 18.1%; Score 90.4; DB 14; Length 717;
Best Local Similarity 70.3%; Pred. No. 3.1e-18;
Matches 135; Conservative 0; Mismatches 56; Indels 1; Gaps 1;
QY 221 TTAGTGAATGCTGTGAGTGAAGGCTTCTGTGGGACAGTAGAAATCGAGGCGATG 280
222 TAATTTGAAGGTGCGGGCGAGATGCTGTGGGCGATTAAAGGCGATAC 281
QY 281 GCCGTTTGAAGTGAATGATGAG-CTACCCCAAACGACAGATTGTCACTCAGTCCGCAAT 339
282 ACCACTTAGTAATCTAATGAAAGCCTATGTGACAGCAGGGTTGTCAATGAGGCGAGAT 341
QY 340 CCGTTTGAAGTGTGATGGGCGAGCGGACAGTGAACAGACACCTACTCACTGTTGAGCA 399
342 CAGATTCCGATTTGAGCGGCGAGCCAACTCAATGAAAGACACACCTGCAAGTTGAGAT 401
QY 400 AGGATGAGAA 411
DB 402 GGAAGATGAGA 413

RESULT 9
BTU89439 1151 bp mRNA linear MAM 10-JAN-2000
LOCUS BTU89439
DEFINITION Bos taurus ubiquitin-like protein mRNA, complete cDNA.
ACCESSION U89439
VERSION U89439.1 GI:1888535
KEYWORDS Bos taurus (cow)
SOURCE Bos taurus
ORGANISM Bos taurus
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovinae; Bos.
REFERENCE 1 (bases 1 to 1151)
AUTHORS Qi, F., Ridpath, J.F. and Berry, E.S.
TITLE Insertion of a bovine SM7B gene in NS4B and duplication of NS3 in
a bovine viral diarrhea virus genome correlate with the
cytopathogenicity of the virus
JOURNAL MEDLINE
MEDLINE 99049626
PUBMED 9833880
REFERENCE 2 (bases 1 to 1151)
AUTHORS Qi, F. and Berry, E.S.
TITLE Direct Submission
JOURNAL Submitted (12-FEB-1997) Oral Biology, University of Alabama at
Birmingham, 250 LHRB 1919 7th Ave. S, Birmingham, AL 35294, USA


```

FEATURES
  source          Location/Qualifiers
                  1. 1151
                  /organism="Bos taurus"
                  /mol_type="mRNA"
                  /db_xref="taxon:9913"
                  /tissue_type="endometrium"
                  548..835
                  /function="unknown"
                  /codon_start=1
                  /product="ubiquitin-like protein"
                  /protein_id="BAB49682.1"
                  /db_xref="gi:1888536"
                  /translation="MADEKPKGVETENNHNILKVAGODGSVVOFKIKRHPISKLM
                  KAYCRQSLSMQIRFRDGPQINTDIPQALEMDEDTIDVFOQQTGVV"

ORIGIN
Query Match      18.1%; Score 90.4; DB 4; Length 1151;
Best Local Similarity 70.3%; Pred. No. 3.3e-18;
Matches 135; Conservative 0; Mismatches 56; Indels 1; Gaps 1;

QY 221 TTATTTGAATGAGTGTGAGTACGAGGTGTTCTGTGGACAGTGAATAATGAGAGGCATCT 280
DB 601 TATTTGAAGTGTGGCGGCGGACAGATGTTCTGTGGTGAAGATTAAAGATTAACGATAC 660
QY 281 GGCGTTAGTGAACGTGATGAGG-CTACCCCAACAGACAGAGATTGTCACTCAGCGCCAT 339
DB 661 ACCACTTGAATGACATGATAGAACCTATTGTGAGACAGCGGTTTGTCAATGAGCAGAT 720
QY 340 CCGATTTCGAGTTGATGAGCGACGCCGACAGTGAACAGACACACCTACTCACTTGGAGGA 399
DB 721 CAGATTCGATTTGACGGCGACGACATCAATGAACAGACACACCTGACAGTTGAGAT 780
QY 400 AGGATGAGAAC 411
DB 781 GGAAATGAAGA 792

RESULT 10
AC127104      211409 bp      DNA      linear      HNG 21-SEP-2002
LOCUS
DEFINITION
Rattus norvegicus clone CH230-128A21, *** SEQUENCING IN PROGRESS
***, 3 unordered pieces.
AC127104.3 GI:23264301
AC127104.3 GI:23264301
HNG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
KEYWORDS
Rattus norvegicus (Norway rat)
SOURCE
ORGANISM
Rattus
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
Bukaryota; Metazoa; Sciuromorphi; Muridae; Murinae;
Rattus

REFERENCE
1 (bases 1 to 211409)
Wuzny,D,Matte, Metzker,M, Lee, Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavares, I., Cesari, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Diya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gibrageorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunnarsson, P., Haaland, W., Hamli, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladik, S. L., Hodgson, A., Hognes, M.,
Hollins, B., Howells, S., Hu, J., Hume, J., Idelbradt, J., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpthy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
REFERENCE
AUTHORS
TITLE
JOURNAL

```

Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, Y., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewa, L., Louised, H., Lozato, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartine, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Ming, A., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Nguyen, N., Nguyen, N., Norris, S., Nwokeleneh, O., Okunolu, G., Olarnunpaso, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfennoch, C., Plopper, F., Polndexter, A., Popovic, D., Primus, E., Pu, L., L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C. D., Smaje, D., Sneed, A., Sodergren, E., Song, X. Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wooden, H., Morley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, J., Zhou, S., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

Direct Submission
Unpublished
2 (bases 1 to 211409)
Morley, K. C.
Direct Submission
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 211409)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (21-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Sep 21, 2002 this sequence version replaced gi:21203531.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). As a result, the sequence may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GNMJ
Center clone name: CH230-128A21
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 187810 bases at least Q40
Consensus quality: 190349 bases at least Q20
Consensus quality: 191813 bases at least Q20
Estimated insert size: 209849; sum-of-contigs estimation
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_dirct_data.html).
NOTE: This is a "working draft" sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence

LOCUS	FEATURES	ORIGIN
CG683622	<p>----- Project Information -----</p> <p>Center project name: GNOF</p> <p>Center clone name: CH230-21D6</p> <p>----- Summary Statistics -----</p> <p>Assembly program: Phrap; version 0.990329</p> <p>Consensus quality: 229402 bases at least Q40</p> <p>Consensus quality: 231763 bases at least Q30</p> <p>Consensus quality: 233057 bases at least Q20</p> <p>Estimated insert size: 260790; sum-of-configs estimation</p> <p>Quality coverage: 3x in Q20 bases; sum-of-configs estimation</p> <p>-----</p> <p>* NOTE: Estimated insert size may differ from sequence length</p> <p>* (see http://www.bgsc.bcm.tmc.edu/docs/Genbankdraft.data.html)</p> <p>* NOTE: This sequence may represent more than one clone.</p> <p>* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.</p> <p>* 1 238985: contig of 238985 bp in length</p> <p>* 238986 239085: gap of unknown length</p> <p>* 239086 240731: contig of 1646 bp in length</p> <p>* 240732 240831: gap of unknown length</p> <p>* 240832 241990: contig of 1159 bp in length.</p> <p>Location/Qualifiers</p> <p>1..241990</p> <p>/organism="Rattus norvegicus"</p> <p>/mol_type="genomic DNA"</p> <p>/db_xref="taxon:10116"</p> <p>/clone="CH230-21D6"</p> <p>1323..2205</p> <p>/note="clone boundary</p> <p>clone end:Sp6</p> <p>site:EcORI</p> <p>end_sequence=BH269859"</p> <p>complement(236297..2377124)</p> <p>/note="clone boundary</p> <p>clone end:T7</p> <p>site:EcORI</p> <p>end_sequence=BH269858"</p>	<p>Query Match 18.0%; Score 90.2; DB 2; Length 241990;</p> <p>Best Local Similarity 62.3%; Pred. No. 9,6e-18;</p> <p>Matches 175; Conservative 0; Mismatches 103; Indels 3; Gaps 2;</p>
Db	221 TTAGTGAATGCTGTGGAGTAGGGGCTCTGTGGGACAGTGAATAATGAGAGGATGT 280	
QY	101717 TAATTGGAAGCGGGTGGACAGAGATGCTCTGTGGGACAGTGAATAATGAGGATAC 101776	
Db	281 GCGGTTTAGTGAAGTGTATGAAG-CTACCCCAAGCAGAGATGTCTGTCAGGCGCAAT 339	
QY	101777 ACCACTTATTAACATATGAAAGCTATTGTGAAACGCGAGGGTTGTCTATGAGCAGAT 101836	
Db	340 CCGTTTCGATTGATGGCGACCGGACAGGTGACAGACACCTACTCAGTTGAGGAG 399	
QY	101837 CAGATTCCGATTGATGGGGAACCAATCAACGAAACAGACACCTGCAACGTTGGAAT 101896	
Db	400 AAGATGAGAACATATGCCAGCAGAGGATGAGACACCTGACAGGCGCAAGGCCCTTAACAC 459	
QY	101897 GAGAGATGAAGATACAGATTGATGTG--TTCCGAGAGAGACAGTGGTGTCTACTAAAAA 101954	
Db	460 ACACACCTTACACCTCACTTGAACAAAGCTGCCAAAGCAGCA 500	
QY	101955 GGGAACTGCTACTTACTTCACAGAGTTTGTATTAGGCAA 101995	

DEFINITION	SEQUENCE 28548 from Patent WO02070737.
ACCESSION	CG683622
VERSION	CG683622.1
KEYWORDS	GI:42206548
SOURCE	
ORGANISM	Homo sapiens (human)
REFERENCE	Homo sapiens
AUTHORS	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
JOURNAL	
FEATURES	1 Location/Qualifiers
source	1..322 /organism="Homo sapiens" /mol_type="unassigned DNA" /db_xref="taxon:9606"
ORIGIN	
Query Match	17.8%; Score 88.8; DB 6; Length 322;
Best Local Similarity	72.3%; Pred. No. 9.1e-18;
Matches	128; Conservative 0; Mismatches 48; Indels 1; Gaps 1;
Qy	221 TTAGTTGAATGCTGTGCAATGATGAGTGGTGTCTGTGTGGCAAGTAAATCGAGAGCGCAT 280
Db	89 TAAATTGAAGGTGGCGGGCGAGATGCTGTGTGTGCATTTAAGTTAAGAGCGCATAC 148
Qy	281 GCCGTTTGTGAACGTATGTAAG-CTACCCCAACGACAGAGATTGTCACTACAGCCCAT 339
Db	149 ACCACTTGTATTAATTAATTAAGCTTAATGAGGACAGAGGATGTCAATGAGGCGCAT 208
Qy	340 CCGTTTCGATTTGATGGGCGACCGCGACATGAGACACACACACCTACTCACTGTTGA 396
Db	209 GAGATTCCGATTTGACGGGCAACCAATCAATGAAACAGACACACCTCGACATTTGA 265
RESULT 13	
LOCUS	BC083326 933 bp mRNA linear ROD 12-OCT-2004
DEFINITION	Mus musculus SMT3 suppressor of mif two 3 homolog 2 (yeast), mRNA (cDNA clone MGC:101988 IMAGE:6819700), complete cds.
ACCESSION	BC083326
VERSION	BC083326
KEYWORDS	MGc.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
	1 (bases 1 to 933)
	Straussberg, R.L., Reingold, E.A., Grouse, L.H., Derge, J.G.,
	Klausner, R.D., Collins, F.S., Wagner, L., Shemen, C.M., Schuler, G.D.,
	Altschul, S.F., Zeeberg, B., Buencow, K.H., Scheffer, C.F., Bat, N.K.,
	Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
	Datchenko, L., Marcina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
	Shapleton, M., Soares, M.B., Bonaldi, M.F., Casavant, T.L.,
	Scheetz, T.E., Brownstein, M.J., Uedon, T.B., Toshitaki, S.,
	Garni, P., Prange, C., Rahn, S.S., Loquellano, N.A., Peters, G.J.,
	Adams, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J.,
	McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
	Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hui, S.W.,
	Vallilon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
	Fahy, J., Helton, E., Ketteman, M., Madan, A., Young, A.C., Shevchenko, Y.,
	Sanchez, A., Whiting, M., Madan, A., Touchman, J.W., Green, E.D.,
	Bouford, G.G., Blakesley, R.W., Grimwood, J., Schmitt, J., Myers, R.M.,
	Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmitt, J., Myers, R.M.,
	Butcherfield, Y.S., Krzyzinski, M.I., Skalska, U., Small, D.E.,
	Schneer, A., Schein, J.E., Jones, S.J., and Marra, M.A.
	Generation and initial analysis of more than 15,000 full-length
	human and mouse cDNA sequences
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16699-16903 (2002)
PUBMED	12477932
REFERENCE	2 (bases 1 to 933)

[illegible]

331 GGAGGATGAAGATACGATTGATGTG--TTCC

331 GGAGGATGAAGATACGATTGATGTG--TTCC

QY 460 ACACACCTACCACTCTGACAAAGCTGCCAAGACCAA 500
DB 399 GGAAGCCTGCTACTTACTCCAGAAATTTGTTATAGACCAA 429

RESULT 15

RATSM3A 987 bp mRNA linear ROD 10-DEC-2001
LOCUS Rat1us norvegicus MIF2 suppressor (SMT3) mRNA, complete cds.
DEFINITION L79949
ACCESSION L79949.1 GI:17467394
VERSION MIF2 suppressor.
KEYWORDS Rat1us norvegicus (Norway rat)
SOURCE Rat1us norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
Rattus.
1 (bases 1 to 987)

REFERENCE Mannen,H., Tsot,S.C., Krushkal,J.S., Li,W.H. and Li,S.S.
AUTHORS The cDNA cloning and molecular evolution of reptile and pigeon
TITLE lactate dehydrogenase isozymes
JOURNAL Mol. Biol. Evol. 14 (11), 1081-1087 (1997)
MEDLINE 98031373
PUBMED 9364765

FEATURES
source Location/Qualifiers
1..987
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/sex="male"
/tissue_type="testis"
1..987
/gene="SMT3"
1..107
/gene="SMT3"
/note="putative"
108..395
/gene="SMT3"
/note="putative"
/codon_start=1
/product="MIF2 suppressor"
/protein_id="AA140175.1"
/db_xref="GI:17467395"
/translation="MADEKPKBGVKTENNHNILKVGADGSVQPKIRHTPLSKLM
KAYCERQGLSMRQIRFRFDQPINETDTPAQLEMEDEDTIVFQQGTGVY"
393..987
/gene="SMT3"
/note="putative"
973..978
/gene="SMT3"
/note="putative"

polyA_signal
ORIGIN
Query Match 17.7%; Score 88.6; DB 10; Length 987;
Best Local Similarity 61.9%; Pred. No. 1,3e-17;
Matches 174; Conservative 0; Mismatches 104; Indels 3; Gaps 2;

QY 221 TTAGTGAATGCTGTGAGTAGGGTGTCTGTGGACAGTAGAAAATCGAGAGCATGT 280
DB 161 TAAATTGAAGTGCGCGGACAGAGTGCTGTGTGTCAGATTAAAGTTAAGAGCATAC 220
QY 281 GCCGTTAGTGAATCTGTGAAG-CTACCCCAACGACAGATGTGCACTCAGGCCAAT 339
DB 221 ACCACTTAGTAACTAATGAAGCTAATGTGAACGCGAGGGTTGTCAATGAGGAGAT 280
QY 340 CCGTTTGAAGTTGATGGCAGCCGACAGTGAACAGACACCTACTCAGTTGAGAGA 399
DB 281 CAGATTCCGTTGATGGGCAACCACTCAAGAAACAGACACACTGCAAGTTGAGAAAT 340
QY 400 AGGATGAGAACAATGGCCAGCAGGATTTGAGAGACCTGA CAGGCGCAGAGCCCTTAAC 459
DB 341 GGAGATGAAGATATGATGATGTG--TTCACAGCAGACAGAGAGGCGTCTACTGAAAA 398

QY 460 ACACACCTACCACTCTTACAAAGCTGCCAAGACCAA 500
DB 399 GGAAGCCTGCTACTTACTCCAGAAATTTGTTATAGACCAA 439

Search completed: February 10, 2005, 14:09:22
Job time : 2412.96 secs

GenCore version 5.1.6
 Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 20:12:02 ; Search time 340.95 Seconds
 (without alignment)
 8681.254 Million cell updates/sec

Title: US-09-926-375b-7_COPY_1_500
 Perfect score: 500
 Sequence: 1 tcgagagatctctctgcgc.....acaagctgcgaagaccaa 500

Scoring table: IDENTITY NUC
 Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : N_Geneseq_16dec04:*
 1: geneseq119808:*
 2: geneseq119908:*
 3: geneseq120008:*
 4: geneseq120018:*
 5: geneseq120018:*
 6: geneseq120028:*
 7: geneseq120028:*
 8: geneseq120038:*
 9: geneseq120038:*
 10: geneseq120038:*
 11: geneseq120038:*
 12: geneseq120048:*
 13: geneseq120048:*

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	500	100.0	17732	3 AAC68300	AAC68300 Lama2/APP
2	500	100.0	20623	3 AAC68294	AAC68294 Lama2/APP
3	88.8	17.8	557	12 ADQ18843	Adq18843 Human sof
4	88.2	17.6	300	10 ADD49509	Add49509 Human lun
5	88.2	17.6	350	6 ABK45405	ABK45405 cDNA enco
6	88.2	17.6	490	6 ABQ58908	ABQ58908 Human col
7	88.2	17.6	561	6 ABK54968	ABK54968 Human col
8	88.2	17.6	628	6 ABK44920	ABK44920 cDNA enco
9	88.2	17.6	644	5 AAS70174	AAS70174 DNA enco
10	88.2	17.6	692	13 ACN40272	ACN40272 Tumour-as
11	88.2	17.6	774	10 ADE84834	Ades84834 Farnesyl
12	88.2	17.6	774	12 ADP82047	ADP82047 Human sen
13	88.2	17.6	1000	5 ABV25085	ABV25085 Human pro
14	88.2	17.6	1000	12 ADP03040	ADP03040 Human hou
15	88.2	17.6	1000	13 ADS88538	ADS88538 Human hou
16	88.2	17.6	1032	6 ABO54780	ABO54780 Human ova
17	88.2	17.6	1039	13 ACN41069	ACN41069 Tumour-as
18	88.2	17.6	1385	12 ADO23237	Ado23237 Human sof
19	86.6	17.3	426	5 AAS81221	AAS81221 DNA enco
20	84.8	17.0	444	3 AAF16398	AAf16398 Human pro

21	83.4	16.7	413	9 ACH46437	ACH46437 Human inf
22	83.4	16.7	516	13 ACN38634	ACN38634 Tumour-as
23	82.8	16.6	383	8 ABX41318	ABx41318 Bovine BS
24	80.8	16.2	277	4 AAI22368	AAI22368 Probe #12
25	80.8	16.2	277	4 ABA67443	ABa67443 Human foe
26	80.8	16.2	277	4 AAI47658	AAI47658 Probe #16
27	80.8	16.2	277	4 ABA49532	ABa49532 Human bre
28	80.8	16.2	277	4 ABA34522	ABa34522 Probe #12
29	80.8	16.2	277	4 AAK41616	AAk41616 Human bon
30	80.8	16.2	277	4 AAK15870	AAk15870 Human bra
31	80.8	16.2	277	4 ABA541207	ABa541207 Human liv
32	80.8	16.2	277	5 AAI08052	AAI08052 Probe #80
33	80.8	16.2	277	6 ABA15625	ABa15625 Human gen
34	79	15.8	1859	3 AAC69094	AAC69094 Human sec
35	79	15.8	1859	3 AAF15591	AAf15591 Human pro
36	77.4	15.5	315	3 AAZ94114	AAz94114 Hematopo
37	73.4	14.7	508	13 ACN38479	ACN38479 Tumour-as
38	73.4	14.7	1037	12 ADQ83862	ADQ83862 Human tum
39	73.4	14.7	1037	12 ADQ87307	ADQ87307 Human tum
40	73.4	14.7	1037	13 ADO85021	ADO85021 Human tum
41	73.4	14.7	1037	13 ACN38478	ACN38478 Tumour-as
42	73.4	14.7	1037	13 ACN38564	ACN38564 Tumour-as
43	73.4	14.7	3903	4 AAF28781	AAf28781 Human GAB
44	73.4	14.7	16862	2 AAX58060	AAX58060 Genomic D
45	73.4	14.7	134292	8 ACA64895	ACA64895 Human GAB

ALIGNMENTS

RESULT 1	AAC68300	standard; DNA; 17732 BP.
ID	AAC68300	standard; DNA; 17732 BP.
XX		
AC	AAC68300;	
XX		
DT	15-SEP-2003 (revised)	
DT	20-FEB-2001 (first entry)	
XX		
DE	Lama2/APPa plasmid coding sequence.	
XX		
KM	Transgenic animal; salivary protein; phytase; phosphorus; animal growth;	
KW	environmental pollution; pig; ds.	
XX		
OS	Mus musculus.	
OS	Escherichia coli.	
OS	Chimeric.	
XX		
PN	WO200064247-A1.	
XX		
PD	02-NOV-2000.	
XX		
PF	20-APR-2000; 2000WO-CA000430.	
XX		
PR	23-APR-1999; 99US-0130508P.	
XX		
PA	(UYGU-) UNIV GDELPH.	
XX		
PI	Forsberg CW, Golovan S, Phillips JP;	
XX		
DR	WPI: 2000-687245/67.	
XX		
PT	P-PSDB; AAB36263.	
XX		
PT	Transgenic non-human animal for gastrointestinal tract specific	
PT	expression of a protein, preferably phytase, comprises a nucleic acid	
PT	sequence including a heterologous transgene construct encoding the	
PT	protein.	
XX		
PS	Claim 14; Fig 23; 152p; English.	
XX		
CC	The present invention provides transgenic animals which produce desired	
CC	proteins, in this case pigs which express phytase in the salivary	
CC	gland. Low phytase production levels result in phytate in the diet being	

CC excreted and causing phosphorus contamination in water, as well as
CC reducing the growth of animals. The invention provides a number of
CC transgenes containing the E. coli APPA phytase coding sequence. (Updated
CC on 15-SEP-2003 to standardise OS field)
XX

Sequence 17732 BP; 4719 A; 4125 C; 4168 G; 4719 T; 0 U; 1 Other;

Query Match 100.0%; Score 500; DB 3; Length 17732;

Best Local Similarity 100.0%; Pred. No. 1,3e-154;

Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 TCGAGAGTATCTTTGTGACGTGCTGCTCCAAAGAGGGGTACTGTTGCCCATAGAAAG 60
    |||
DB 1 TCGAGAGTATCTTTGTGACGTGCTGCTCCAAAGAGGGGTACTGTTGCCCATAGAAAG 60
QY 61 ATCTAACTAATTAATTAATCCCTCACCCGCAATCTTTCAGTCACTAAGTAGACGAT 120
    |||
DB 61 ATCTAACTAATTAATTAATCCCTCACCCGCAATCTTTCAGTCACTAAGTAGACGAT 120
QY 121 TGTGAAACAAGTTCTCCAAAGAGAGATACAGATGAGTCCGTATAGGGTGGACCTGGCTG 180
    |||
DB 121 TGTGAAACAAGTTCTCCAAAGAGAGATACAGATGAGTCCGTATAGGGTGGACCTGGCTG 180
QY 181 CTGAGAGACACCTGATCTGACTAAGAGAGCAGCGTGTAGTTGAATGGTGTGAGT 240
    |||
DB 181 CTGAGAGACACCTGATCTGACTAAGAGAGCAGCGTGTAGTTGAATGGTGTGAGT 240
QY 241 AGGGTGGTCTGTGGAGCAGTAGAAAATCGAAGGCGATGCGCTTTAGTGAATGATG 300
    |||
DB 241 AGGGTGGTCTGTGGAGCAGTAGAAAATCGAAGGCGATGCGCTTTAGTGAATGATG 300
QY 301 AAGCTACCCCAACGACAGAGATTGTCAGTCAAGGCCAATCCGTTTGAATTTGAGGGCA 360
    |||
DB 301 AAGCTACCCCAACGACAGAGATTGTCAGTCAAGGCCAATCCGTTTGAATTTGAGGGCA 360
QY 361 GCCGACAGTGAAGACAGACACCTACTCAGTTGAGGAAGATGAGAACAAATGGCCAGC 420
    |||
DB 361 GCCGACAGTGAAGACAGACACCTACTCAGTTGAGGAAGATGAGAACAAATGGCCAGC 420
QY 421 AGGGATTGAGAGACCTTGAAGGCGCAAGGCCCTTAACAACAACACTTCACTCTTG 480
    |||
DB 421 AGGGATTGAGAGACCTTGAAGGCGCAAGGCCCTTAAACAACAACACTTCACTCTTG 480
QY 481 ACAAAAGCTGCCAAAGACCA 500
    |||
DB 481 ACAAAAGCTGCCAAAGACCA 500

RESULT 2
AAC68294 standard; DNA; 20623 BP.
XX
AC AAC68294;
XX
DT 15-SEP-2003 (revised)
DT 20-FEB-2001 (first entry)
XX
XX Lama2/APPA plasmid coding sequence.
XX
DE Transgenic animal; salivary protein; phytase; phosphorus; animal growth;
XX
KM environmental pollution; pig; ds.
XX
XX Mus musculus.
XX
OS Escherichia coli.
XX
OS Chimeric.
XX
XX WO200064247-A1.
XX
PN 02-NOV-2000.
XX
PD 20-APR-2000; 2000WO-CA000430.
XX
PF 23-APR-1999; 99US-0130508P.
XX
PR
```

XX (UYGU-) UNIV GUELPH.
XX
XX Forsberg CW, Golovan S, Phillips JP;
XX
XX WPI; 2000-687245/67.
XX
XX P-PSDB; AAB36257.

Transgenic non-human animal for gastrointestinal tract specific
PT expression of a protein, preferably phytase, comprises a nucleic acid
PT sequence including a heterologous transgene construct encoding the
PT protein.

PS Claim 56; Fig 5; 152pp; English.

XX The present invention provides transgenic animals which produce desired
XX proteins, in this case pigs which expresses phytase in the salivary
XX gland. Low phytase production levels result in phytate in the diet being
XX excreted and causing phosphorus contamination in water, as well as
XX reducing the growth of animals. The invention provides a number of
XX transgenes containing the E. coli APPA phytase coding sequence. (Updated
XX on 15-SEP-2003 to standardise OS field)

Sequence 20623 BP; 5449 A; 4847 C; 4902 G; 5424 T; 0 U; 1 Other;

Query Match 100.0%; Score 500; DB 3; Length 20623;

Best Local Similarity 100.0%; Pred. No. 1,4e-154;

Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 TCGAGAGTATCTTTGTGACGTGCTGCTCCAAAGAGGGGTACTGTTGCCCATAGAAAG 60
    |||
DB 1 TCGAGAGTATCTTTGTGACGTGCTGCTCCAAAGAGGGGTACTGTTGCCCATAGAAAG 60
QY 61 ATCTAACTAATTAATTAATCCCTCACCCGCAATCTTTCAGTCACTAAGTAGACGAT 120
    |||
DB 61 ATCTAACTAATTAATTAATCCCTCACCCGCAATCTTTCAGTCACTAAGTAGACGAT 120
QY 121 TGTGAAACAAGTTCTCCAAAGAGAGATACAGATGAGTCCGTATAGGGTGGACCTGGCTG 180
    |||
DB 121 TGTGAAACAAGTTCTCCAAAGAGAGATACAGATGAGTCCGTATAGGGTGGACCTGGCTG 180
QY 181 CTGAGAGACACCTGATCTGACTAAGAGAGCAGCGTGTAGTTGAATGGTGTGAGT 240
    |||
DB 181 CTGAGAGACACCTGATCTGACTAAGAGAGCAGCGTGTAGTTGAATGGTGTGAGT 240
QY 241 AGGGTGGTCTGTGGAGCAGTAGAAAATCGAAGGCGATGCGCTTTAGTGAATGATG 300
    |||
DB 241 AGGGTGGTCTGTGGAGCAGTAGAAAATCGAAGGCGATGCGCTTTAGTGAATGATG 300
QY 301 AAGCTACCCCAACGACAGAGATTGTCAGTCAAGGCCAATCCGTTTGAATTTGAGGGCA 360
    |||
DB 301 AAGCTACCCCAACGACAGAGATTGTCAGTCAAGGCCAATCCGTTTGAATTTGAGGGCA 360
QY 361 GCCGACAGTGAAGACAGACACCTACTCAGTTGAGGAAGATGAGAACAAATGGCCAGC 420
    |||
DB 361 GCCGACAGTGAAGACAGACACCTACTCAGTTGAGGAAGATGAGAACAAATGGCCAGC 420
QY 421 AGGGATTGAGAGACCTTGAAGGCGCAAGGCCCTTAAACAACAACACTTCACTCTTG 480
    |||
DB 421 AGGGATTGAGAGACCTTGAAGGCGCAAGGCCCTTAAACAACAACACTTCACTCTTG 480
QY 481 ACAAAAGCTGCCAAAGACCA 500
    |||
DB 481 ACAAAAGCTGCCAAAGACCA 500
```

RESULT 3

ADQ18843

ID ADQ18843 standard; DNA; 557 BP.

XX ADQ18843;

DT 26-AUG-2004 (first entry)

XX DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 1662.
XX KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
XX KW de.
XX OS Homo sapiens.
XX PN MO2004048938-A2.
XX PD 10-JUN-2004.
XX PF 26-NOV-2003; 2003MO-US038193.
XX PR 26-NOV-2002; 2002US-0429739P.
XX PA (PROT-) PROTEIN DESIGN LABS INC.
XX PI Aziz N, Ginsburg WM, Zlotnik A;
XX DR WPI; 2004-441208/41.
XX PT Early detection of soft tissue sarcoma comprises determining expression
XX of a gene in a first soft tissue sample and a normal soft tissue sample
XX and comparing the gene expression, also useful in treating soft tissue
XX sarcoma.
XX PS Example 2; SEQ ID NO 1662; 210pp; English.
XX CC The invention relates to a novel method for detecting soft tissue sarcoma
XX CC which comprises obtaining a first soft tissue sample from an individual
XX CC and a normal soft tissue sample from the same or different individual,
XX CC determining the expression of a gene in both samples and comparing the
XX CC expression of the gene in the first soft tissue sample, where a higher level
XX CC of protein expression in the first soft tissue sample indicates the
XX CC presence of soft tissue sarcoma. The method of the invention has
XX CC cytostatic applications and may be useful for detecting soft tissue
XX CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
XX CC acid sequences may be useful in diagnostic and screening applications.
XX CC The current sequence is that of a human soft tissue sarcoma-upregulated
XX CC DNA of the invention. The current sequence is not shown within the
XX CC specification per se but was submitted in CD format by the inventor.
SQ Sequence 557 BP; 150 A; 136 C; 130 G; 140 T; 0 U; 1 Other;
Query Match 17.8%; Score 88.8; DB 12; Length 557;
Best Local Similarity 72.3%; Pred. No. 1.3e-18;
Matches 128; Conservative 0; Mismatches 48; Indels 1; Gaps 1;
QY 221 TTGATTGAATGGTGTGAGTAGTGGTGTCTGTGGACAGTAAATCCAGAGGCTAT 280
DB 211 TAAATTGAGGTGGCGGAGGAGATGTTCTGTGGTCAAGTTTAAGTAGAGGCTAC 270
QY 281 GCGGTTAGTGAAGTGAAG-CTACCCCAACGACAGAGATTGTCAGTCAAGCCAAAT 339
DB 271 ACCACTTAATTAATGAAGGCTATTGTGAACGACAGGATTGTCATAGAGCGCAT 330
QY 340 CCGTTTGAAGTTGATGGAGCCGCGACAGTGAACAGACACCTCACTCAGTTGA 396
DB 331 CAGATTCCGATTGACGGGCAACCAATCAATGAACAGACACCTCAGAGGTTGA 387
RESULT 4
ADD49509/c
ID ADD49509 standard; cDNA; 300 BP.
AC ADD49509;
XX 15-JAN-2004 (first entry)
XX Human lung cancer associated cDNA 61571931.
XX DE Human lung cancer associated cDNA 61571931.
XX KW Human; se; lung cancer antigen; cytostatic; lung cancer; gene therapy;

KM vaccine; T-cell; tumour.
XX OS Homo sapiens.
XX PN US2003194764-A1.
XX PD 16-OCT-2003.
XX PF 04-APR-2002; 2002US-00116712.
XX PR 05-APR-2001; 2001US-0282289P.
XX PR 05-OCT-2001; 2001US-0327511P.
XX PA (CORI-) CORIXA CORP.
XX PI Bangur CS, Switzer A;
XX DR WPI; 2003-844452/78.
XX PT New isolated polypeptides and polynucleotides useful for diagnosing,
XX preventing and treating cancer, particularly lung cancer.
XX PS Claim 1; SEQ ID NO 241; 250pp; English.
XX CC The invention relates to an isolated polynucleotide (a) comprising any of
XX CC the 666 fully defined nucleotide sequences appearing as ADD49269 -
XX CC ADD49935, ADD49936 and ADD49938, complements of (a); sequences of at
XX CC least 20 contiguous residues of (a); sequences that hybridise to (a)
XX CC under highly stringent conditions; sequences having at least 75 or 90%
XX CC identity to (a); or degenerate variants of (a). Also included are an
XX CC isolated polypeptide (b) (comprising: sequences encoded by the new
XX CC polynucleotide; any of the 4 amino acid sequences fully defined in the
XX CC specification; or sequences having at least 70 or 90% identity to the
XX CC sequence in (a) or (b)), an expression vector comprising the above
XX CC polynucleotide operably linked to an expression control sequence, a host
XX CC cell transformed or transfected with the above expression vector, an
XX CC isolated antibody, or its antigen-binding fragment, that specifically
XX CC binds to the above polypeptide, an oligonucleotide that hybridises to the
XX CC above-mentioned nucleotide sequences under highly stringent conditions, a
XX CC fusion protein comprising at least one polypeptide cited above, detecting
XX CC the presence of a cancer in a patient (comprising: obtaining a biological
XX CC sample from the patient; contacting the biological sample with a binding
XX CC agent that binds to the polypeptide, or with the oligonucleotide cited
XX CC above; detecting in the sample an amount of the polypeptide that binds to
XX CC the binding agent, or an amount of a polynucleotide that hybridises to
XX CC the oligonucleotide; and comparing the amount of polypeptide, or
XX CC polynucleotide that hybridises to the oligonucleotide, to a predetermined
XX CC cut-off value and then determining the presence of a cancer in the
XX CC patient), a method for stimulating and/or expanding T-cells specific for
XX CC a tumour protein (comprising contacting T-cells with the above
XX CC polypeptide, polynucleotide or antigen-presenting cells that express the
XX CC polynucleotide, under conditions and for a time sufficient to permit the
XX CC stimulation and/or expansion of T-cells), an isolated T-cell population
XX CC comprising T-cells prepared by the method, a composition comprising a
XX CC first component selected from physiological carriers and
XX CC immunostimulants, and a second component selected from the above
XX CC polypeptide, polynucleotide, antibody, fusion protein, T-cell population
XX CC and antigen-presenting cells that express the above polypeptide,
XX CC stimulating an immune response in a patient (comprising administering to
XX CC the patient the above composition) treating lung cancer in a patient
XX CC (comprising administering to the patient the above composition and a
XX CC diagnostic kit (comprising: at least one oligonucleotide cited above; or
XX CC at least one antibody cited above and a detection reagent, where the
XX CC detection reagent comprises a reporter group). The composition and
XX CC methods are useful in diagnosing, preventing and treating cancer,
XX CC particularly lung cancer. The present sequence is a lung cancer-
XX CC associated antigen cDNA of the invention.
SQ Sequence 300 BP; 73 A; 77 C; 54 G; 96 T; 0 U; 0 Other;
Query Match 17.6%; Score 88.2; DB 10; Length 300;
Best Local Similarity 72.3%; Pred. No. 1.5e-18;
Matches 128; Conservative 0; Mismatches 48; Indels 1; Gaps 1;

QY 221 TTAGTTGAATGCTGTGAGTAGGCTGTTCTGTGGGACAGTAAATCGAAGCCATGT 280
DB 274 TAATTTGAAGGTGGCGGGGACAGATGCTTGTGTGTGAGTTAAATTAAGGACATAC 215
QY 261 GCCGTTTGTGTAACATGATGAGAG-CTACCCCAAGACAGAGATTGTGAGTCCAGGCCAT 339
DB 214 ACCACTTGTAACTAAATGAAGCCTATTGTGAAAGCAGAGGATTTGTCAATGAGGCAAT 155
QY 340 CCGTTTTCAGTTTGTATGGGACCGGACAGTGAAGACACACCTACTCAGTTTGA 396
DB 154 CAGATTCGATTTTGAACGGGACCAATCAATGAAGAAACAGACACCTGCACAGTTGA 98

RESULT 5
ABK45405
ID ABK45405 standard; cDNA; 350 BP.

AC ABK45405;
XX
XX 05-JUN-2002 (first entry)
XX
XX
XX cDNA encoding colon tumour protein, SEQ ID No 956.
XX
XX Human; colon tumour; vaccine; colon cancer; immunogenic; immunotherapy;
XX gene; ss.
XX
XX Homo sapiens.
XX
XX WO200212328-A2.
XX
XX 14-FEB-2002.
XX
XX 31-JUL-2001; 2001WO-US024218.
XX
XX 03-AUG-2000; 2000US-0223283P.
XX PR 28-MAR-2001; 2001US-0279763P.
XX PR 29-JUN-2001; 2001US-0302051P.
XX
XX (CORI-) CORIXA CORP.
XX
XX King GE, Meagher MJ, Xu J, Secrist H;
XX
XX WPI; 2002-241739/29.
XX
XX New colon cancer polypeptides and polynucleotides, useful as vaccines,
XX for diagnosing, preventing, and treating colon cancer, and as markers for
XX the progression of cancer.
XX
XX Claim 1; SEQ ID NO 956; 147bp; English.

XX The invention relates to polynucleotides encoding colon tumour proteins,
XX CC The polynucleotides and encoded polypeptides are useful in pharmaceutical
XX compositions, such as vaccines, for the diagnosis, prevention, and
XX treatment of colon cancer. Polynucleotide sequences may be used as
XX hybridisation probes or primers, and in the design and preparation of
XX ribozyme molecules for inhibiting expression of tumour polypeptides and
XX proteins in tumour cells. The compositions are useful for stimulating an
XX immune response against cancer, particularly for the immunotherapy of
XX colon cancer, and as markers for the progression of cancer. ABK4450-
XX ABK6237 represent coding sequences of human colon tumour proteins of the
XX invention. Note: With the exception of SEQ ID No 1 and 2, the sequence
XX data for this patent did not form part of the printed specification but
XX was supplied by the European Patent Office
XX
XX

XX Sequence 350 BP; 110 A; 69 C; 94 G; 77 T; 0 U; 0 Other;

Query Match 17.6%; Score 88.2; DB 6; Length 350;
Best Local Similarity 72.3%; Pred. No. 1.6e-18;
Matches 128; Conservative 0; Mismatches 48; Indels 1; Gaps 1;

QY 221 TTAGTTGAATGCTGTGAGTAGGCTTCTGTGGGACAGTAAATCGAAGCCATGT 280
DB 274 TAATTTGAAGGTGGCGGGGACAGATGCTTGTGTGTGAGTTAAATTAAGGACATAC 215

DB 77 TAATTTGAAGGTGGCGGGGACAGATGCTTGTGTGAGTTAAATTAAGGACATAC 136
QY 261 GCCGTTTGTGTAACATGATGAGAG-CTACCCCAAGACAGAGATTGTGAGTCCAGGCCAT 339
DB 137 ACCACTTGTAACTAAATGAAGCCTATTGTGAAAGCAGAGGATTTGTCAATGAGGCAAT 196
QY 340 CCGTTTTCAGTTTGTATGGGACCGGACAGTGAAGACACACCTACTCAGTTTGA 396
DB 197 CAGATTCGATTTTGAACGGGACCAATCAATGAAGAAACAGACACCTGCACAGTTGA 253

RESULT 6
ABO58908/C
ID ABO58908 standard; cDNA; 490 BP.

XX ABO58908;
XX
XX 02-AUG-2002 (first entry)
XX
XX
XX Human colon cancer related nucleotide sequence SEQ ID No:2603.
XX
XX
XX Human; colon cancer; cancer; tissue profiling; forensic; mapping;
XX genetic analysis; diagnostic; antisense therapy; gene; ss.
XX
XX Homo sapiens.
XX
XX WO200229086-A2.
XX
XX 11-APR-2002.
XX
XX 02-OCT-2001; 2001WO-US030732.
XX
XX 02-OCT-2000; 2000US-0237271P.
XX
XX (FARB) BAYER CORP.
XX
XX Burgess C, Aetle JH, Carroll E, Catino TV, Dwiwedi P, Molino GA;
XX Thiglinglam A, Lewis ME;
XX
XX WPI; 2002-426115/45.
XX
XX New isolated nucleic acid that is differentially expressed in cancer
XX tissues useful for determining the presence of colon cancer in a cell or
XX tissue type, and in antisense therapy.
XX
XX Claim 1; Fig 1; 796bp; English.

XX ABO56306 to ABO60787 represent isolated nucleic acids (I) differentially
XX expressed in cancer tissues. ABB78993 to ABB79004 represent proteins
XX encoded by the ABO60776 to ABO60787 nucleic acid sequences. (I) can be
XX used in antisense therapy. An antibody immunoreactive with a polypeptide
XX encoded by (I) is useful for detecting cancer in a patient sample, and
XX for detecting the presence or absence of a polynucleotide encoded by a
XX nucleic acid which hybridises to (I) in a cell. A probe/primer derived
XX from (I) can be used for determining the presence of a nucleic acid which
XX hybridises to (I), and for determining the phenotype of cells in a sample
XX of cells from a patient. (I) is useful for determining the presence of
XX colon cancer in a cell or tissue type, for determining the presence or
XX state of other type of cancer, in antisense therapy, to generate
XX macroarrays on a solid surface, to identify a chromosome on which the
XX corresponding gene resides, and in tissue profiling, forensics, genetic
XX analysis, mapping and diagnostic applications. (I) can be used to raise
XX antibodies, and to screen for peptide analogues and antagonists
XX
XX

XX Sequence 490 BP; 126 A; 110 C; 109 G; 145 T; 0 U; 0 Other;

Query Match 17.6%; Score 88.2; DB 6; Length 490;
Best Local Similarity 72.3%; Pred. No. 1.9e-18;
Matches 128; Conservative 0; Mismatches 48; Indels 1; Gaps 1;

QY 221 TTAGTTGAATGCTGTGAGTAGGCTTCTGTGGGACAGTAAATCGAAGCCATGT 280
DB 393 TAATTTGAAGGTGGCGGGGACAGATGCTTGTGTGAGTTAAATTAAGGACATAC 334

QY 281 GCCGTTTGAAGTGAAG-CTACCCCAACGACAGATTGTCACTCAGGCCAAT 339
DB 333 ACCACTTGAAGCTAATGAAAGCTTGTGTGACGACAGGAGTTGTCAATGAGCAGAT 274
QY 340 CCGTTTGAAGTTGAGGCGAGCCGACAGTGAAGACACACTTCACTAGTTTGA 396
DB 273 CAGATTCCGATTGACGCGACCAATCAATGAAACAGACACTTGCACAGTTTGA 217

RESULT 7

ABK54968
ID ABK54968 standard; cDNA; 561 BP.

AC ABK54968;

DT 18-JUN-2002 (first entry)

DE Human colon cancer-associated cDNA, SEQ ID No 438.

KW Human; colon cancer; immunogenic; vaccine; tumour; gene; ss.

OS Homo sapiens.

PN WO200212280-A2.

PD 14-FEB-2002.

PF 30-JUL-2001; 2001WO-US023826.

PR 03-AUG-2000; 2000US-0223265P.

PR 02-OCT-2000; 2000US-0237406P.

PR 20-MAR-2001; 2001US-0277495P.

PR 03-JUL-2001; 2001US-0302702P.

PA (CORI-) CORIXA CORP.

PI Pyle RA, Xu J, Secrist H;

PT WPI; 2002-257462/30.

PS Novel polynucleotide encoding colon tumor polypeptides, useful as

PS vaccines for treating colon cancers.

PS Claim 1; Page 275; 425pp; English.

CC The invention relates to isolated polynucleotides (I) encoding colon

CC tumour polypeptides (II). (I) is useful for stimulating an immune

CC response in a patient and treating colon cancer in a patient.

CC Oligonucleotides derived from (I) are useful for determining the presence

CC of cancer in a patient. (I) and (II) are useful in pharmaceutical

CC compositions, e.g. vaccines, and other compositions for the diagnosis and

CC treatment of colon cancer. A composition comprising a first component

CC selected from physiologically acceptable carriers and immunostimulants,

CC and an antigen-presenting cell expressing (II) is useful for inhibiting

CC development of cancer in a patient. (I) is useful in the design and

CC preparation of ribozyme molecules for inhibiting expression of tumour

CC polypeptides and (I). ABK54531-ABK55464 represent human colon cancer cDNA

CC sequences of the invention

XX Sequence 561 BP; 154 A; 126 C; 134 G; 147 T; 0 U; 0 Other;

QY 340 CCGTTTGAAGTTGAGGCGAGCCGACAGTGAAGACACACTTCACTAGTTTGA 396
DB 274 CAGATTCCGATTGACGCGACCAATCAATGAAACAGACACTTGCACAGTTTGA 330

RESULT 8

ABK44920
ID ABK44920 standard; cDNA; 628 BP.

AC ABK44920;

DT 05-JUN-2002 (first entry)

DE cDNA encoding colon tumour protein, SEQ ID No 471.

KW Human; colon tumour; vaccine; colon cancer; immunogenic; immunotherapy;

KW gene; ss.

OS Homo sapiens.

PN WO200212328-A2.

PD 14-FEB-2002.

PF 31-JUL-2001; 2001WO-US024218.

PR 03-AUG-2000; 2000US-0223283P.

PR 28-MAR-2001; 2001US-0279763P.

PR 29-JUN-2001; 2001US-0302051P.

PA (CORI-) CORIXA CORP.

PI King GE, Meagher MJ, Xu J, Secrist H;

PT WPI; 2002-241739/29.

PS New colon cancer polypeptides and polynucleotides, useful as vaccines,

PS for diagnosing, preventing, and treating colon cancer, and as markers for

PS the progression of cancer.

PS Claim 1; SEQ ID NO 471; 147pp; English.

CC The invention relates to polynucleotides encoding colon tumour proteins.

CC The polynucleotides and encoded polypeptides are useful in pharmaceutical

CC compositions, such as vaccines, for the diagnosis, prevention, and

CC treatment of colon cancer. Polynucleotide sequences may be used as

CC hybridisation probes or primers, and in the design and preparation of

CC ribozyme molecules for inhibiting expression of tumour polypeptides and

CC proteins in tumour cells. The compositions are useful for stimulating an

CC immune response against cancer, particularly for the immunotherapy of

CC colon cancer, and as markers for the progression of cancer. ABK4450-

CC ABK46237 represent coding sequences of human colon tumour proteins of the

CC invention. Note: With the exception of SEQ ID No 1 and 2, the sequence

CC data for this patent did not form part of the printed specification but

CC was supplied by the European Patent Office

XX Sequence 628 BP; 186 A; 132 C; 136 G; 172 T; 0 U; 2 Other;

XX Query Match 17.6%; Score 88.2; DB 6; Length 628;

XX Best Local Similarity 72.3%; Pred. No. 2; 1e-18;

XX Matches 128; Conservative 0; Mismatches 48; Indels 1; Gaps 1;

QY 221 TTAAGTTGATGCTGAGTGAAGTGTCTGTGTGACAGTGAAGAAATCGAGAGCATGT 280

DB 77 TAAATTTGAAGGTGCGGGGACGATGTTCTGTGTGACAGTTTAAGATTAAGAGGCATAC 136

QY 281 GCCGTTTGAAGTGAAG-CTACCCCAACGACAGATTGTCACTCAGGCCAAT 339

DB 137 ACCACTTGAAGCTAATGAAAGCTTGTGTGACGACAGGAGTTGTCAATGAGGCAGAT 196

QY 340 CCGTTTGAAGTTGAGGCGAGCCGACAGTGAAGACACACTTCACTAGTTTGA 396

DB 197 CAGATTCGATTGACGGGCAACCAATCATGATAAGACAGACACACTGCACAGTTGGA 253

RESULT 9
AAS70174
ID AAS70174 standard; cDNA; 644 BP.
XX
XX AAS70174;
AC AAS70174;
XX
XX 13-FEB-2002 (first entry)
DT
XX
XX DNA encoding novel human diagnostic protein #5978.
DE
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KM food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
OS
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US008631.
XX
XX 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
XX (HYSF-) HYSFQ INC.
PA
XX Dmanac RT, Liu C, Tang YT;
PI
XX WPI; 2001-639362/73.
DR P-PSDB; ABG05987.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 1; SEQ ID NO 5978; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
CC
XX
SQ Sequence 644 BP; 176 A; 139 C; 142 G; 172 T; 0 U; 15 Other;

Query Match 17.6%; Score 88.2; DB 5; Length 644;
Best Local Similarity 72.3%; Pred. No. 2.1e-18;
Matches 128; Conservative 0; Mismatches 48; Indels 1; Gaps 1;

OY 221 TTAGTTGAATGCTGTGAGTGGGTTCTTCTGTGGGAACATGAAATCGAGAGCATGT 280
DB 164 TAAATTTGAAGTGCGGGGAGGATGTTCTGTGTGATTAAGATTAAAGGCAATAC 223

OY 281 GCGGTTAGTGAAGTGTGAG-CTACCCAAACAGACAGATTGTCAGTCAGGCCAAT 339
DB 224 ACCACTAGTAACCTAATGAAGCCTTGTGTGAACAGAGGATGTGATAGGCGAGAT 283

OY 340 CCGTTTCGAGTTTGTATGGGCGGACAGTGAACAGACACCTTACTCAGTTGGA 396
DB 284 CAGATTCGATTGACGGGCAACCAATCATGATAAGACAGACACCTGCACAGTTGGA 340

RESULT 10
ACN40272
ID ACN40272 standard; cDNA; 692 BP.
XX
XX ACN40272;
AC
XX 18-NOV-2004 (first entry)
DT
XX
XX Tumour-associated antigenic target (TAT) cDNA DNA326400. SEQ ID NO:4957.
DE
XX
XX Tumour-associated antigenic target; TAT; human; overexpression; cancer;
KM tumour; diagnosis; cell proliferative disorder; breast cancer;
KM colorectal cancer; lung cancer; ovarian cancer; liver cancer;
KM central nervous system cancer; bladder cancer; pancreatic cancer;
KM cervical cancer; melanoma; leukaemia; hybridisation probe;
KM chromosome identification; chromosome mapping; gene mapping;
KM gene therapy; cytostatic; gene; ss.
XX
XX Homo sapiens.
OS
XX WO2004030615-A2.
XX
XX 15-APR-2004.
XX
XX 29-SEP-2003; 2003WO-US028547.
XX
XX 02-OCT-2002; 2002US-0414971P.
PR
XX (GENTH) GENENTECH INC.
PA
XX Wu TD, Zhang Z, Zhou Y;
PI
XX WPI; 2004-347921/32.
DR
XX
XX New tumor-associated antigenic target polypeptides and nucleic acids,
XX useful in preparing a medicament for treating or detecting a
XX proliferative disorder, e.g. breast, lung, colorectal, ovarian or
XX prostate cancer or tumor.
XX
XX Claim 1; SEQ ID NO 4957; 7273pp; English.

CC The invention relates to human tumour-associated antigenic target (TAT)
CC polypeptides, and their related nucleic acids. The TAT polypeptides are
CC overexpressed in cancer tissues compared to normal tissues, and may thus
CC serve as effective targets for the diagnosis and treatment of cancer in
CC mammals. The invention also relates to the TAT nucleic acids and
CC sequences at least 80% identical to the TAT nucleic acids and
CC polypeptides; expression vectors and host cells comprising a TAT nucleic
CC acid; an antibody specific for a TAT polypeptide; fusion proteins comprising a
CC molecule which binds to a TAT polypeptide; fusion proteins comprising a
CC TAT polypeptide; and methods and compositions for the treatment or
CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
CC antibodies, antagonists, binding molecules and compositions are useful
CC for diagnosing or treating a cell proliferative disorder associated with
CC increased TAT expression, particularly cancers such as breast cancer,
CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
CC cancer, pancreatic cancer, cervical cancer, cancers of the central
CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be
CC used as hybridisation probes, in chromosome and gene mapping, in
CC chromosome identification and in gene therapy. The present sequence
CC represents a TAT nucleic acid of the invention
XX
SQ Sequence 692 BP; 195 A; 143 C; 157 G; 197 T; 0 U; 0 Other;

Query Match 17.6%; Score 88.2; DB 13; Length 692;
Best Local Similarity 72.3%; Pred. No. 2.2e-18;
Matches 128; Conservative 0; Mismatches 48; Indels 1; Gaps 1;
QY 221 TTAGTTGAATGAGTGTGAGTAGGAGTGTCTGTGAGCAGTAGAATAATCGAGGCAATG 280
DB 134 TAATTTGAAGGTGGCGGCGGAGATGTTCTGTGTGCGAGTTTAAGATTAAAGGGCATAC 193
QY 281 GCCGTTTGAATGAACTGATGAGAG-CTACCCCAAACGACAGATTGTCTAGTCAAGCCCAAT 339
DB 194 ACCACTTAGTAATTAATGAAAGCTATTGTGAAACGACAGGAGTTGTCAATGAGGCAAGAT 253
QY 340 CGCTTTCGAGTTTGAATGAGGCGGACCCGACAGTGAAGACACACTACTAGTTTGA 396
DB 254 CAGATTCCGATTGACGGGCAACCAATCAATGAAACAGACACCTGCACAGTTTGA 310
RESULT 11
ADE84834
ID ADE84834 standard; DNA; 774 BP.
XX ADE84834;
AC
XX
DT 29-JAN-2004 (first entry)
DE Farnesyl transferase inhibitor modulated leukemia associated gene #53.
XX
XX 88; cytostatic; farnesyl transferase inhibitor; gene expression;
KM quinolone; leukemia; cancer.
OS Homo sapiens.
XX
XX WO2003038129-A2.
PN
XX 08-MAY-2003.
PD
XX 30-OCT-2002; 2002WO-US034784.
PF
XX 30-OCT-2001; 2001US-0338997P.
PR 30-OCT-2001; 2001US-0340081P.
PR 30-OCT-2001; 2001US-0340938P.
PR 30-OCT-2001; 2001US-0341012P.
XX
XX (ORTH) ORTHO CLINICAL DIAGNOSTICS INC.
PA
PI Rapont M;
XX
DR WPI; 2003-513497/48.
XX
XX Determining whether a patient will respond to treatment with a farnesyl
PT transferase inhibitor, by analyzing the expression of gene that is
PT differentially modulated in the presence of the inhibitor.
XX
XX
PS Disclosure; SEQ ID NO 53; 346bp; English.
XX
XX The invention relates to a method of determining whether a patient will
CC respond to treatment with a farnesyl transferase inhibitor (FTI), by
CC analyzing the expression of gene that is differentially modulated in the
CC presence of an FTI. The method is useful for determining whether a
CC patient will respond to treatment with a FTI such as (B)-6-(amino(4-
CC chlorophenyl)(1-methyl-1H-imidazol-5-yl)methyl)-4-(3-chlorophenyl)-1-
CC methyl-2-(1H)quinolone, monitoring the therapy of a patient, treating a
CC patient with leukemia with FTI if the analysis indicates that the patient
CC will respond. This sequence corresponds to a gene whose expression may be
CC modulated in the presence of FTI.
XX
SQ Sequence 774 BP; 222 A; 166 C; 170 G; 212 T; 0 U; 4 Other;
Query Match 17.6%; Score 88.2; DB 10; Length 774;
Best Local Similarity 72.3%; Pred. No. 2.3e-18;
Matches 128; Conservative 0; Mismatches 48; Indels 1; Gaps 1;
QY 221 TTAGTTGAATGAGTGTGAGTAGGAGTGTCTGTGAGCAGTAGAATAATCGAGGCAATG 280

DB 144 TAATTTGAAGGTGGCGGCGGAGATGTTCTGTGAGCAGTTTAAGATTAAAGGCAATAC 203
QY 281 GCCGTTTGAATGAACTGATGAGAG-CTACCCCAAACGACAGATTGTCTAGTCAAGCCCAAT 339
DB 204 ACCACTTAGTAATTAATGAAAGCTATTGTGAAACGACAGGAGTTGTCAATGAGGCAAGAT 263
QY 340 CGCTTTCGAGTTTGAATGAGGCGGACCCGACAGTGAAGACACACTACTAGTTTGA 396
DB 264 CAGATTCCGATTGACGGGCAACCAATCAATGAAACAGACACCTGCACAGTTTGA 320
RESULT 12
ADP82047
ID ADP82047 standard; DNA; 774 BP.
XX ADP82047;
AC
XX
DT 26-AUG-2004 (first entry)
DE Human sentrin-2 DNA.
XX
XX Sentrin-2; SMT3H2; SMT3A; SUMO-3; hyperproliferative disorder; cancer;
KM cytostatic; gene therapy; human; chromosome 17q25; gene; ds.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH CDS 91..378
FT /*tag= a
FT /product= "Human sentrin-2 protein"
XX
XX US2004110151-A1.
PN
XX 10-JUN-2004.
PD
XX 10-DEC-2002; 2002US-00316638.
PF
XX 10-DEC-2002; 2002US-00316638.
PR
XX 10-DEC-2002; 2002US-00316638.
PR
XX (ISIS-) ISIS PHARM INC.
XX
XX Ward DT, Dobie KW;
PI
XX WPI; 2004-440340/41.
DR P-PSDB; ADP82119.
DR GENBANK; X99585.1.
XX
XX New oligonucleotide compound that inhibits expression of sentrin-2,
PT useful for preparing a composition for treating hyperproliferative
PT disorder, e.g. cancer.
XX
XX
PS Claim 1; SEQ ID NO 4; 35bp; English.
XX
XX The present invention is directed to antisense oligonucleotides targeted
CC to sentrin-2 (also known as SMT3H2, SMT3A and SUMO-3) and which modulate
CC the expression of sentrin-2. The invention is useful for preparing a
CC composition for treating hyperproliferative disorder such as cancer. The
CC invention acts as a cytostatic agent. The invention is also useful in
CC gene therapy. The present sequence is human sentrin-2 DNA. The sentrin-2
CC gene is located on chromosome 17q25.
XX
SQ Sequence 774 BP; 222 A; 166 C; 170 G; 212 T; 0 U; 4 Other;
Query Match 17.6%; Score 88.2; DB 12; Length 774;
Best Local Similarity 72.3%; Pred. No. 2.3e-18;
Matches 128; Conservative 0; Mismatches 48; Indels 1; Gaps 1;
QY 221 TTAGTTGAATGAGTGTGAGTAGGAGTGTCTGTGAGCAGTAGAATAATCGAGGCAATG 280
DB 144 TAATTTGAAGGTGGCGGCGGAGATGTTCTGTGTGCGAGTTTAAGATTAAAGGGCATAC 203
QY 281 GCCGTTTGAATGAACTGATGAGAG-CTACCCCAAACGACAGATTGTCTAGTCAAGCCCAAT 339

Db 204 ACCACTTAGTAACCTAATGAAGCCTATTGTGAACGACGAGGATGTCAATAGGCAGAT 263
 QY 340 CCGTTTCAGATTGATGGGCGACCGGACAGTGAACAGACACCTACTCAGTTGGA 396
 Db 264 CAGATTCCGATTGTGACGGGGCAACCAATCAATGAACAGACACCTGCACAGTTGGA 320

RESULT 13

ABV25085
 ID ABV25085 standard; cDNA; 1000 BP.

ABV25085;

16-SEP-2002 (first entry)

Human prostate expression marker cDNA 25076.

Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker; pharmacogenomic marker; gene; 88.

Homo sapiens.

WO200160860-A2.

23-AUG-2001.

20-FEB-2001; 2001WO-US005171.

17-FEB-2000; 2000US-0183119P.

16-MAR-2000; 2000US-0189862P.

25-MAY-2000; 2000US-0207454P.

09-JUN-2000; 2000US-0211314P.

18-JUL-2000; 2000US-0219007P.

13-DEC-2000; 2000US-0255281P.

(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

Schlegel R, Endege WO, Monahan JE;

WPI; 2001-662795/76.

Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer.

Claim 1; Page 4877; 11750pp; English.

The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (I) is useful for: (a) assessing whether a patient is afflicted with prostate cancer; (b) monitoring the efficacy of a test compound to inhibit prostate cancer in a patient; (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient; (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient; (e) selecting a composition for inhibiting prostate cancer in a patient; (f) assessing the prostate cell carcinogenic potential of a compound; (g) determining whether prostate cancer has metastasized in a patient; (h) assessing the aggressiveness or indolence of prostate cancer in a patient; (i) is also useful as a pharmacodynamic or pharmacogenomic marker

Sequence 1000 BP; 303 A; 186 C; 228 G; 283 T; 0 U; 0 Other;

Query Match 17.6%; Score 88.2; DB 5; Length 1000;

Best Local Similarity 72.3%; Pred. No. 2.6e-18; Mismatches 128; Conservative 0; Indels 1; Gaps 1;

221 TTAGTTGATGAGTGTGAGTAGGCTGTCTGTGGACAGTAAATCGAGGAGATGT 280

72 TAATTTGAAGTGGCGGGCGAGATGTTCTGTGGTCAATTTAAATTAAGGAGATAC 131

281 GCCGTTTGAAGTGAATGAG-CTACCCCAACGACAGAGATTGTCAAGTCAAGCCAAAT 339

Db 132 ACCACTTAGTAACCTAATGAAGCCTATTGTGAACGACGAGGATGTCAATAGGCAGAT 191
 QY 340 CCGTTTCAGATTGATGGGCGACCGGACAGTGAACAGACACCTACTCAGTTGGA 396
 Db 192 CAGATTCCGATTGTGACGGGGCAACCAATCAATGAACAGACACCTGCACAGTTGGA 248

RESULT 14

ADP03040
 ID ADP03040 standard; cDNA; 1000 BP.

ADP03040;

29-JUL-2004 (first entry)

Human housekeeping gene cDNA #81.

ss; gene; human; housekeeping gene; capture probe; DNA microarray; cancer.

Homo sapiens.

JP2004135552-A.

13-MAY-2004.

16-OCT-2002; 2002JP-00302278.

16-OCT-2002; 2002JP-00302278.

(NIGA) NGK INSULATORS LTD.

WPI; 2004-404613/38.

Novel human housekeeping gene commonly expressed in different human tissues, useful as capture probe in DNA microarray.

Claim 1; SEQ ID NO 81; 940pp; Japanese.

The invention relates to a human housekeeping gene. The housekeeping gene is useful as a capture probe in DNA microarray. The housekeeping gene is useful as a reference with respect to the normal gene expression measured on various conditions, and an expression database of normal human tissue is useful as reference information for using a disease related gene for diagnosis or treatment. The housekeeping gene when used as capture probe, enables diagnosis of cancer. The present sequence represents a human housekeeping gene cDNA.

Sequence 1000 BP; 329 A; 184 C; 189 G; 298 T; 0 U; 0 Other;

Query Match 17.6%; Score 88.2; DB 12; Length 1000;

Best Local Similarity 72.3%; Pred. No. 2.6e-18; Mismatches 128; Conservative 0; Indels 1; Gaps 1;

221 TTAGTTGATGAGTGTGAGTAGGCTGTCTGTGGACAGTAAATCGAGGAGATGT 280

124 TAATTTGAAGTGGCGGGCGAGATGTTCTGTGGTCAATTTAAATTAAGGAGATAC 183

281 GCCGTTTGAAGTGAATGAG-CTACCCCAACGACAGAGATTGTCAAGTCAAGCCAAAT 339

184 ACCACTTAGTAACCTAATGAAGCCTATTGTGACAGACAGGAGATTGTCAATAGGCAGAT 243

340 CCGTTTCAGATTGATGGGCGACCGGACAGTGAACAGACACCTACTCAGTTGGA 396

244 CAGATTCCGATTGTGACGGGGCAACCAATCAATGAACAGACACCTGCACAGTTGGA 300

RESULT 15

AD888538
 ID AD888538 standard; cDNA; 1000 BP.

AD888538;

DT 18-NOV-2004 (first entry)
 XX
 DE Human housekeeping gene cDNA sequence SEQ ID NO:81.
 XX
 KW housekeeping gene; human; probe; hybridisation; DNA microarray;
 XX disease investigation; disease diagnosis; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO2004035785-A1.
 XX
 PD 29-APR-2004.
 XX
 PF 16-OCT-2002; 2002WO-JP010753.
 XX
 PR 16-OCT-2002; 2002WO-JP010753.
 XX
 PA (NICA) NGK INSULATORS LTD.
 XX
 PI Aburatani H, Yamamoto S;
 XX
 DR WPI; 2004-357219/33.
 XX
 PT Human housekeeping and tissue-specific genes, gene sets and transcription
 PT products, probes and microarrays derived from them for reference use in
 PT investigation and diagnosis of disease.
 PS
 XX
 PS Claim 5; SEQ ID NO 81; 372pp; Japanese.
 XX
 CC The present invention describes 1189 housekeeping genes and gene sets
 CC containing them, expressed in 35 different human tissues. Also described:
 CC (1) RNA and cDNA transcription products of the housekeeping genes; (2)
 CC oligonucleotide probes hybridising to the housekeeping genes; (3) DNA
 CC microarrays containing the oligonucleotide probes; (4) 1704 genes
 CC expressed specifically in particular tissues; (5) RNA and cDNA
 CC transcription products of the tissue-specific genes; (6) oligonucleotide
 CC probes hybridising to the tissue-specific genes; and (7) DNA microarrays
 CC containing the oligonucleotide probes. The tissue-specific genes are
 CC expressed specifically in one of brain, thymus, pituitary, thyroid,
 CC trachea, lung, breast, skin, skeletal muscle, heart, liver, spleen,
 CC kidney, adrenal gland, pancreas, stomach, small intestine, colon,
 CC bladder, prostate gland, testis, ovary, placenta, uterus, bone marrow,
 CC foetal brain or foetal liver. The housekeeping genes can be used as
 CC reference genes for the investigation and diagnosis of disease. The
 CC present sequence represents a human housekeeping gene cDNA sequence which
 CC is given in the exemplification of the present invention.
 CC
 SQ Sequence 1000 BP; 329 A; 184 C; 189 G; 298 T; 0 U; 0 Other;
 Query Match 17.6%; Score 88.2; DB 13; Length 1000;
 Best Local Similarity 72.3%; Pred. No. 2.6e-18;
 Matches 128; Conservative 0; Mismatches 48; Indels 1; Gaps 1;
 QY 221 TTAGTTGAATGCTGTGAGTGGTGTCTGTGGACATGAAATCGAGAGCATGT 280
 DB 124 TAAATTTGAAGTGGCGGGCGGAGATGCTGTGTGGATTAAGATTAAAGGCAATAC 183
 QY 281 GCCGTTTAGTGAATGATGAG-CTAACCAGACAGAGATTGTCACTCAGGCCAAT 339
 DB 184 ACCACTTAGTAACTAATGAAAGCTTATTTGAAACGACAGGATTGTCAATGAGGCAGAT 243
 QY 340 CCGTTTGCAGTTTATGGGCGAGCCGACAGTGAACAGACACACTCACTCAGTTGA 396
 DB 244 CAGATTTCGATTTCAGCGGCAACCAATCAATGAACAGACACACTGCACAGTTGGA 300

Search completed: February 10, 2005, 09:40:59
 Job time : 344.95 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 22:08:58 ; Search time 103.713 Seconds
(without alignments)
7888.487 Million cell updates/sec

Title: US-09-926-375B-7_COPY_1_500

Perfect score: 500
Sequence: 1 tcgagagatctcttcgcagc.....acaagctccaaagaccaa 500

Scoring table: IDENTITY NUC
Gapop 10_0, Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Issued Patence NA: *
1: /cgn2_6/prodata/1/ina/5A COMB.seq: *
2: /cgn2_6/prodata/1/ina/5B COMB.seq: *
3: /cgn2_6/prodata/1/ina/6A COMB.seq: *
4: /cgn2_6/prodata/1/ina/6B COMB.seq: *
5: /cgn2_6/prodata/1/ina/PCUS COMB.seq: *
6: /cgn2_6/prodata/1/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	88.2	17.6	754	US-09-949-016-1577	Sequence 3577, Ap
2	77.7	15.4	84587	US-09-949-016-15733	Sequence 15733, A
3	73.4	14.7	15652	US-09-422-936-60	Sequence 60, Appl
4	72.4	14.5	601	US-09-949-016-148687	Sequence 148687, A
5	72.4	14.5	124480	US-09-949-016-15921	Sequence 15921, A
6	71.4	14.3	473	US-08-853-974-2	Sequence 2, Appl1
7	71.4	14.3	473	US-09-172-988-2	Sequence 2, Appl1
8	71.4	14.3	1728	US-09-949-016-5228	Sequence 5228, Ap
9	53.2	10.6	7218	US-08-232-463-14	Sequence 14, Appl
10	45.6	9.1	18970	US-09-949-016-15319	Sequence 15319, A
11	45.2	9.0	87350	US-08-781-891-79	Sequence 79, Appl
12	45.2	9.0	87350	US-09-618-166-79	Sequence 79, Appl
13	45.2	9.0	87543	US-09-791-211-3	Sequence 3, Appl1
14	45.2	9.0	143776	US-09-949-001-39	Sequence 29, Appl
15	45.2	9.0	144034	US-09-949-001-35	Sequence 35, Appl
16	41.8	8.4	28843	US-09-949-016-17325	Sequence 17325, A
17	41.8	8.4	914	US-09-270-767-14655	Sequence 14655, A
18	40.2	8.0	16365	US-09-949-016-16970	Sequence 16970, A
19	38.6	7.7	601	US-09-949-016-91168	Sequence 91168, A
20	38.6	7.7	175236	US-09-949-016-14353	Sequence 14353, A
21	36.2	7.4	96202	US-09-949-016-16433	Sequence 16433, A
22	36.2	7.2	32172	US-09-949-016-16440	Sequence 16440, A
23	35.8	7.2	601	US-09-949-016-182341	Sequence 182341, A
24	35.8	7.2	601	US-09-949-016-182342	Sequence 182342, A
25	35.8	7.2	2990	US-09-009-816-19	Sequence 19, Appl
26	34.6	6.9	34088	US-09-949-016-14449	Sequence 14449, A
27	34.6	6.9	95255	US-09-949-016-17067	Sequence 17067, A

28	34.2	6.8	589	US-09-949-016-186352	Sequence 186352, A
29	34.2	6.8	260247	US-09-949-016-13358	Sequence 13358, A
30	34	6.8	105210	US-09-949-016-14158	Sequence 14158, A
31	33.8	6.8	1118	US-09-799-451-694	Sequence 694, Appl
32	33.8	6.8	786431	US-09-751-389-3	Sequence 3, Appl1
33	33	6.6	601	US-09-949-016-23475	Sequence 23475, A
34	33	6.6	601	US-09-949-016-161297	Sequence 161297, A
35	33	6.6	601	US-09-949-016-185917	Sequence 185917, A
36	33	6.6	45571	US-09-949-016-16262	Sequence 16262, A
37	33	6.6	99370	US-09-949-016-12816	Sequence 12816, A
38	33	6.6	99370	US-09-949-016-17540	Sequence 17540, A
39	33	6.6	124110	US-09-949-016-13353	Sequence 13353, A
40	33	6.6	331814	US-09-949-016-12008	Sequence 12008, A
41	33	6.6	331814	US-09-949-016-17056	Sequence 17056, A
42	32.8	6.6	33152	US-09-949-016-16719	Sequence 16719, A
43	32.6	6.5	181	US-09-513-999C-24911	Sequence 24911, A
44	32.6	6.5	19732	US-09-949-016-14923	Sequence 14923, A
45	32.6	6.5	168174	US-10-071-411A-63	Sequence 63, Appl

ALIGNMENTS

RESULT 1
US-09-949-016-3577
Sequence 3577, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: C10010307
CURRENT APPLICATION NUMBER: US/09/949, 016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241, 755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237, 768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231, 498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3577
LENGTH: 754
TYPE: DNA
ORGANISM: Human
US-09-949-016-3577

Query Match 17.6%; Score 88.2; DB 4; Length 754;
Best Local Similarity 72.3%; Pred. No. 1.1e-19;
Matches 128; Conservative 0; Mismatches 48; Indels 1; Gaps 1;

QY	221	TTAGTTGATGATGCTGTGAGTAGGCTTCTGTGGACAGTAGAATAATCGAGGCGATGT	280
DB	140	TAAATTGACAGTGGCGGCGGAGATGCTTGTGTGACATTTAAGTTAAGAGGCTAC	199
QY	281	GCCGTTGATGAACTGATGAG-CTRACCCAAACAGACAGATGTCACTAGGCCAAT	339
DB	200	ACCACTTGAATTAATTAATGAAAGCTAATGTGACAGCAGAGATTTGTCATGAGGCGAT	259
QY	340	CCGTTTCGATTTGATGAGGCGGCGGACAGTGAAGACACACCTACTCAGTTGGA	396
DB	260	CAGATTCCGATTGACGCGGACCAATCAATGAAACAGACACTGCACAGTTGGA	316

RESULT 2
US-09-949-016-15733/C
Sequence 15733, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

```
/ FILE REFERENCE: CL001307
/ CURRENT APPLICATION NUMBER: US/09/949,016
/ CURRENT FILING DATE: 2000-04-14
/ PRIOR APPLICATION NUMBER: 60/241,755
/ PRIOR FILING DATE: 2000-10-20
/ PRIOR APPLICATION NUMBER: 60/237,768
/ PRIOR FILING DATE: 2000-10-03
/ PRIOR APPLICATION NUMBER: 60/231,498
/ PRIOR FILING DATE: 2000-09-08
/ NUMBER OF SEQ ID NOS: 207012
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 15733
/ LENGTH: 84587
/ TYPE: DNA
/ ORGANISM: Human
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(84587)
/ OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15733

Query Match          15.4%; Score 77; DB 4; Length 84587;
Best Local Similarity 68.4%; Pred. No. 9,1e-15;
Matches 121; Conservative 0; Mismatches 55; Indels 1; Gaps 1;

QY 221 TTAGTTGAATGCTGTGAGTGGGCTTCTGTGGACATGAAAATCGAGAGCATGT 280
DB 21875 TAATTTGAAGTGGCAGGCGAGGAGGTTCTGTGTGCTTGAATTAAGGCGCATAC 21816
QY 281 GCGGTTAGTGAACGTATGGAAG-CTACCCAAAGCAGAGATTGTGCTAGCGCCAAAT 339
DB 21815 ACCACTTGAATTAATGAAAGCCTGTTGTGAACATGAGGATGATGACACAGAT 21756
QY 340 CCGTTTCGAGTTGATGGGACCGGACAGTGAACACACACCTACTCAGTTGGA 396
DB 21755 CAGATTCGATTTGACAGGCAACCATCATGAAAGACACACCTGCTCAGTTGGA 21639

RESULT 3
US-09-422-936-60
/ Sequence 60, Application US/09422936
/ Patent No. 6465213
/ GENERAL INFORMATION:
/ APPLICANT: Ekstrand, Jonas
/ TITLE OF INVENTION: NEW NUCLEOTIDE SEQUENCES
/ FILE REFERENCE: 06275-165002
/ CURRENT APPLICATION NUMBER: US/09/422,936
/ CURRENT FILING DATE: 1999-10-22
/ PRIOR APPLICATION NUMBER: US 09/242,608
/ PRIOR FILING DATE: 1999-02-19
/ PRIOR APPLICATION NUMBER: PCT/SE98/01947
/ PRIOR FILING DATE: 1998-10-27
/ PRIOR APPLICATION NUMBER: SWEDEN 9703914-2
/ PRIOR FILING DATE: 1997-10-27
/ PRIOR APPLICATION NUMBER: SWEDEN 9800864-2
/ PRIOR FILING DATE: 1998-03-16
/ PRIOR APPLICATION NUMBER: SWEDEN 9802575-2
/ PRIOR FILING DATE: 1998-07-17
/ NUMBER OF SEQ ID NOS: 85
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 60
/ LENGTH: 15652
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(15652)
/ OTHER INFORMATION: n = A, T, G, or C
/ NAME/KEY: misc_feature
/ LOCATION: (1)...(15652)
/ OTHER INFORMATION: r = G or A; y = T/U or C; m = A or C;
/ OTHER INFORMATION: k = G or T/U; s = G or C; w = A or T/U;
/ OTHER INFORMATION: b = G, C, or T/U; d = A, G, or T/U;
```

```
/ OTHER INFORMATION: h = A, C, or T/U; v = A, G, or C
/ NAME/KEY: exon
/ LOCATION: (3419)...(3444)
/ NAME/KEY: intron
/ LOCATION: (3445)...(3908)
/ NAME/KEY: exon
/ LOCATION: (3909)...(3993)
/ NAME/KEY: intron
/ LOCATION: (3994)...(4694)
/ NAME/KEY: exon
/ LOCATION: (4695)...(4898)
/ NAME/KEY: intron
/ LOCATION: (4899)...(5652)
/ NAME/KEY: exon
/ LOCATION: (5653)...(5838)
/ NAME/KEY: intron
/ LOCATION: (5839)...(7184)
/ NAME/KEY: exon
/ LOCATION: (7185)...(7205)
/ NAME/KEY: intron
/ LOCATION: (7206)...(8310)
/ NAME/KEY: exon
/ LOCATION: (8311)...(8806)
/ NAME/KEY: intron
/ LOCATION: (8807)...(12271)
/ NAME/KEY: exon
/ LOCATION: (12272)...(12406)
/ NAME/KEY: intron
/ LOCATION: (12407)...(12820)
/ NAME/KEY: exon
/ LOCATION: (12821)...(12991)
/ NAME/KEY: intron
/ LOCATION: (12992)...(14089)
/ NAME/KEY: exon
/ LOCATION: (14090)...(14191)
/ NAME/KEY: intron
/ LOCATION: (14192)...(14477)
/ NAME/KEY: exon
/ LOCATION: (14478)...(14543)
/ NAME/KEY: intron
/ LOCATION: (14544)...(15002)
/ NAME/KEY: exon
/ LOCATION: (15003)...(15194)
/ NAME/KEY: intron
/ LOCATION: (15195)...(15652)
US-09-422-936-60

Query Match          14.7%; Score 73.4; DB 3; Length 15652;
Best Local Similarity 71.2%; Pred. No. 6,4e-14;
Matches 126; Conservative 0; Mismatches 46; Indels 5; Gaps 2;

QY 221 TTAGTTGAATGCTGTGAGTGGGCTTCTGTGGACAGTAAATCGAGAGCATGT 280
DB 9 TAATTTGAAGTGGGCGGCGAGATGTTCTGTGTGCTTGAATTAAGGCGCATAC 68
QY 281 GCGGTTAGTGAACGTATGGAAG-CTACCCAAAGCAGAGATTGTGCTAGCGCCAAAT 339
DB 69 ACCACTTGAATTAATGAAAGCCTATTTGTGAACAGAGGATGTCAATGAGCGAAT 128
QY 340 CCGTTTCGAGTTGATGGGACCGGACAGTGAACACACACCTACTCAGTTGGA 396
DB 129 CAGATTCGATTTGACAGGCGCAAC----AATGAAACAGACACCTGCGACAGTTGGA 181

RESULT 4
US-09-949-016-148687/c
/ Sequence 148687, Application US/09949016
/ Patent No. 6812339
/ GENERAL INFORMATION:
/ APPLICANT: VENTER, J. Craig et al.
/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
/ WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
/ FILE REFERENCE: CL001307
```

CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 148687
LENGTH: 601
TYPE: DNA
ORGANISM: Human
US-09-949-016-148687

Query Match 14.5%; Score 72.4; DB 4; Length 601;
Best Local Similarity 70.2%; Pred. No. 2.5e-14;
Matches 125; Conservative 0; Mismatches 51; Indels 2; Gaps 2;

QY 221 TTAGTTGATGATGAGTGGAGTGGTGTCTGTGGACAGTAAATCGAGAGCATGT 280
DB 228 TAAATTTGAAGGTGGTGGGAGGGTGTCTGTGTGTCAGTTAAGTTAAGAGGCATAC 169
QY 281 GCCGTTTAGTGAAGTGTGAG-CTACCCCAACGACAGAGATTGTCACTCAGGCCAAT 339
DB 168 ACCACTTAGTAACTAATGAAAGCTTATGTGAATGACAGGATTTGTCATGAGGCAGAT 109
QY 340 CCGTTTCAGTTTGTATGAGGCGGCGGA-CAGTGAGACAGACACACTACTCAGTTGGA 396
DB 108 CAGACTGTGATTTGACCGGCAACCAATCAATGAAAAAGACACCTGTACAGTTGGA 51

RESULT 5
US-09-949-016-15921/c

Sequence 15921, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: C1001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 15921
LENGTH: 124480
TYPE: DNA
ORGANISM: Human
US-09-949-016-15921

Query Match 14.5%; Score 72.4; DB 4; Length 124480;
Best Local Similarity 70.2%; Pred. No. 4.2e-13;
Matches 125; Conservative 0; Mismatches 51; Indels 2; Gaps 2;

QY 221 TTAGTTGATGATGAGTGGAGTGGTGTCTGTGGACAGTAAATCGAGAGCATGT 280
DB 59936 TAAATTTGAAGGTGGGAGGGTGTCTGTGTGTCAGTTAAGTTAAGAGGCATAC 59877
QY 281 GCCGTTTAGTGAAGTGTGAG-CTACCCCAACGACAGAGATTGTCACTCAGGCCAAT 339
DB 59876 ACCACTTAGTAACTAATGAAAGCTTATGTGAATGACAGGATTTGTCATGAGGCAGAT 59817
QY 340 CCGTTTCAGTTTGTATGAGGCGGCGGA-CAGTGAGACAGACACACTACTCAGTTGGA 396
DB 59816 CAGACTGTGATTTGACCGGCAACCAATCAATGAAAAAGACACCTGTACAGTTGGA 59759

RESULT 6
US-08-853-974-2

Sequence 2, Application US/08853974
Patent No. 5840534
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
TITLE OF INVENTION: HUMAN SMT3-LIKE PROTEIN
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/853,974
FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0289 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 473 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: LUNGFET03
CLONE: 2361410
US-08-853-974-2

Query Match 14.3%; Score 71.4; DB 2; Length 473;
Best Local Similarity 67.1%; Pred. No. 4.8e-14;
Matches 116; Conservative 0; Mismatches 56; Indels 1; Gaps 1;

QY 226 TGAATGCTGTGAGTGGTGTCTGTGGACAGTAAATCGAGAGCATGTGCGCT 285
DB 139 TGAAGTGGCGGGGAGGAGCGCTCGTGTGCAAGTTCAAGATCAAGAGCACACCGCGC 198
QY 286 TTACTGACTGATGAGAG-CTACCCCAACGACAGAGATTGTCACTCAGGCCAATCCGT 344
DB 199 TGAGCAAGCTGATGAGAGGCTTACTGCGAGGAGGAGGCGCTTGTCAATGAGGCAAGTCAGAT 258
QY 345 TCGAGTTGATGGGAGCGGAGCGAGAGTGAAGACACACCTACTCAGTTGAG 397
DB 259 TCAGGTTGACGGGACGCCAATCAATGAACTGACACTCCAGCACAGCTGGAG 311

RESULT 7
US-09-172-988-2

Sequence 2, Application US/09172988
Patent No. 6136955
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
TITLE OF INVENTION: HUMAN SMT3-LIKE PROTEIN
NUMBER OF SEQUENCES: 3

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/172,988
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/853,974
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0289 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-845-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 473 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: LUNGFET03
CLONE: 2361410
US-09-172-988-2

Query Match      14.3%; Score 71.4; DB 3; Length 473;
Best Local Similarity 67.1%; Pred. No. 4.8e-14;
Matches 116; Conservative 0; Mismatches 56; Indels 1; Gaps 1;

QY      226 TGAATGCTGTGAGTGGGTGTTCTGTGGACAGTAAATCGAGGCGATGCGCT 285
DB      139 TGAAGTGGCGCGGCGAGACGCTCGGTGTGCAAGTCAAGACAGAGCAGCCGC 198
QY      286 TTAGTAACTGATGAAG-CTACCCAAACAGACAGATTGTAGTCAGGCCAATCCGT 344
DB      199 TGAGCAAGCTGATGAAGGCTTACTGCGAGAGCGAGGCTTGTCAATGAAGCAGATCAGAT 258
QY      345 TCGAGTTGATGGCGAGCGCGACAGTGAAGCAGACACCTACTCAGTTGGAG 397
DB      259 TCAGGTTGACGGGCGAGCCATCATGAATGAAGTGAAGCTCCAGCAGAGCTGGAG 311

RESULT 8
US-09-949-016-5228
Sequence 5228, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: C1001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5228
```

```

LENGTH: 1728
TYPE: DNA
ORGANISM: Human
US-09-949-016-5228

Query Match      14.3%; Score 71.4; DB 4; Length 1728;
Best Local Similarity 67.1%; Pred. No. 9.6e-14;
Matches 116; Conservative 0; Mismatches 56; Indels 1; Gaps 1;

QY      226 TGAATGCTGTGAGTGGGTGTTCTGTGGACAGTAAATCGAGGCGATGCGCT 285
DB      146 TGAAGTGGCGCGGCGAGACGCTCGGTGTGCAAGTCAAGAGGACACGCTGC 205
QY      286 TTAGTAACTGATGAAG-CTACCCAAACAGACAGATTGTAGTCAGGCCAATCCGT 344
DB      206 TGAGCAAGCTGATGAAGGCTTACTGCGAGAGCGAGGCTTGTCAATGAAGCAGATCAGAT 265
QY      345 TCGAGTTGATGGCGAGCGCGACAGTGAAGCAGACACCTACTCAGTTGGAG 397
DB      266 TCAGGTTGACGGGCGAGCCATCATGAATGAAGTGAAGCTCCAGCAGAGCTGGAG 318

RESULT 9
US-08-232-463-14/C
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-9300
TELEFAX: (703) 683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: PTZ9Pt-F1s
US-08-232-463-14

Query Match      10.6%; Score 53.2; DB 1; Length 7218;
Best Local Similarity 8.5%; Pred. No. 3.5e-07;
```



```

CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/618,166
FILING DATE: 17-Jul-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: McMaisters, David D.
REGISTRATION NUMBER: 33,963
REFERENCE/DOCKET NUMBER: 240052.419C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 79:
SEQUENCE CHARACTERISTICS:
LENGTH: 87350 base pairs
TYPE: nucleic acid
STRADEDNESS: single
TOPOLOGY: Linear
SEQUENCE DESCRIPTION: SEQ ID NO: 79:
US-09-618-166-79

Query Match          9.0%; Score 45.2; DB 4; Length 87350;
Best Local Similarity 66.3%; Pred. No. 0.00072;
Matches 79; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

QY 282 CCGTTAGTGAAGTGAAG-CTACCCCAACGACAGATGTGTCAGTCAGGCCAATC 340
DB 12102 CCGCTTAGTAACCAATGAAGGCTATGTAAAGGGGACGGTTGCCAATGAGGACAGATC 12043
QY 341 CGTTTCAGTTGATGGGCGCCGCGACAGTGAACAGACACACTACTACTAGTTGGAGG 398
DB 12042 AGGTTCAATTGACAGGCGACCAATCAATGAACAGACACATCTGCACTTTGGAAG 11985

RESULT 13
US-09-791-211-3/c
Sequence 3, Application US/09791211
Patent No. 6448080
GENERAL INFORMATION:
APPLICANT: Donna T. Ward
TITLE OF INVENTION: ANTISENSE MODULATION OF WRN EXPRESSION
FILE REFERENCE: RTS-0205
CURRENT APPLICATION NUMBER: US/09/791,211
CURRENT FILING DATE: 2001-02-23
NUMBER OF SEQ ID NOS: 90
SEQ ID NO 3
LENGTH: 87543
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: unsure
LOCATION: 7421
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 7427
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 11609
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 12605
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 12742
```

```

OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 29370
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 29422
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 29979
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 29980
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 29981
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 30136
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 30140
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 31205
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 31206
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 31592
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 31095
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 31160
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 34066
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 34072
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 36816
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 39020
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 42164
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 42459
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 46808
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 46823
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 46826
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 47291
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 52786
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 52787
OTHER INFORMATION: unknown
```



```

NAME/KEY: unsure
LOCATION: 53384
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 54684
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 59215
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 59235
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 59242
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 63290
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 66614
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 68660
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 68697
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 68718
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 68733
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 68739
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 69785
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 79134
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 79198
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 86336
OTHER INFORMATION: unknown
US-09-791-211-3

```

```

Query Match      9.0%; Score 45.2; DB 3; Length 87543;
Best Local Similarity 66.9%; Pred. No. 0.00073;
Matches 79; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

QY      282 CCGTTAGTGAAGTGAAG-CTACCCCAACGACAGAGATTGTGAGTCCAGGCCAATC 340
DB      12295 CCGTTAGTGAAGTGAAG-CTACCCCAACGACAGAGATTGTGAGTCCAGGCCAATC 12236
QY      341 CGTTTCAGTTTGTGGGCGGCGGACAGAGACACACTACTCAGTTGGAGG 398
DB      12235 AGTTTCAATTGACAGGCGGCGGACAGAGACACACTACTGCACTTTGGAGG 12178

```

```

RESULT 14
US-09-949-001-29/C
Sequence 29, Application US/09949001
Patent No. 6825336
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH OSTEOPOROSIS, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: C1000789

```

```

CURRENT APPLICATION NUMBER: US/09/949,001
CURRENT FILING DATE: 2003-01-15
PRIOR APPLICATION NUMBER: 60/231,323
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 848
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 29
LENGTH: 143776
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(143776)
OTHER INFORMATION: n = A,T,C or G
US-09-949-001-29

```

```

Query Match      9.0%; Score 45.2; DB 4; Length 143776;
Best Local Similarity 66.9%; Pred. No. 0.00094;
Matches 79; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

QY      282 CCGTTAGTGAAGTGAAG-CTACCCCAACGACAGAGATTGTGAGTCCAGGCCAATC 340
DB      99899 CCGTTAGTGAAGTGAAG-CTACCCCAACGACAGAGATTGTGAGTCCAGGCCAATC 99840
QY      341 CGTTTCAGTTTGTGGGCGGCGGACAGAGACACACTACTCAGTTGGAGG 398
DB      99839 AGTTTCAATTGACAGGCGGCGGACAGAGACACACTACTGCACTTTGGAGG 99782

```

```

RESULT 15
US-09-949-001-35/C
Sequence 35, Application US/09949001
Patent No. 6825336
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH OSTEOPOROSIS, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: C1000789
CURRENT FILING DATE: 2003-01-15
PRIOR APPLICATION NUMBER: 60/231,323
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 848
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 35
LENGTH: 144034
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(144034)
OTHER INFORMATION: n = A,T,C or G
US-09-949-001-35

```

```

Query Match      9.0%; Score 45.2; DB 4; Length 144034;
Best Local Similarity 66.9%; Pred. No. 0.00094;
Matches 79; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

QY      282 CCGTTAGTGAAGTGAAG-CTACCCCAACGACAGAGATTGTGAGTCCAGGCCAATC 340
DB      99899 CCGTTAGTGAAGTGAAG-CTACCCCAACGACAGAGATTGTGAGTCCAGGCCAATC 99840
QY      341 CGTTTCAGTTTGTGGGCGGCGGACAGAGACACACTACTCAGTTGGAGG 398
DB      99839 AGTTTCAATTGACAGGCGGCGGACAGAGACACACTACTGCACTTTGGAGG 99782

```

Search completed: February 10, 2005, 17:42:45
Job time : 107.713 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 22:13:58 ; Search time 364.334 Seconds
(Without alignments)
8097.594 Million cell updates/sec

Title: US-09-926-375B-7_COPY_1_500

Perfect score: 500

Sequence: 1 tcgagagatccttgcagc.....acaagctccaaagaccacaa 500

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 537673 seqs, 2950229984 residues 10757346

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
20: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
21: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
22: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	88.8	17.8	322	17	US-10-242-535A-28548
2	88.8	17.8	322	17	US-10-085-783A-28548
3	88.8	17.8	557	18	US-10-723-860-1662
4	88.2	17.6	300	16	US-10-116-712-241
5	88.2	17.6	303	17	US-10-242-535A-6748
6	88.2	17.6	303	17	US-10-085-783A-6748
7	88.2	17.6	339	17	US-10-057-475B-10573
8	88.2	17.6	339	17	US-10-154-884B-10573
9	88.2	17.6	350	9	US-09-920-300A-956
10	88.2	17.6	350	13	US-10-033-528-956
11	88.2	17.6	350	16	US-10-099-926-956

12	88.2	17.6	461	17	US-10-242-535A-41493	Sequence 41493, A
13	88.2	17.6	461	17	US-10-085-783A-41493	Sequence 41493, A
14	88.2	17.6	465	17	US-10-242-535A-39663	Sequence 39663, A
15	88.2	17.6	465	17	US-10-085-783A-39663	Sequence 39663, A
16	88.2	17.6	490	11	US-09-969-034-2603	Sequence 2603, A
17	88.2	17.6	561	9	US-09-919-580-438	Sequence 438, App
18	88.2	17.6	628	9	US-09-920-300A-471	Sequence 471, App
19	88.2	17.6	628	13	US-10-033-528-471	Sequence 471, App
20	88.2	17.6	628	16	US-10-099-926-471	Sequence 471, App
21	88.2	17.6	774	18	US-10-316-638-4	Sequence 4, App
22	88.2	17.6	774	18	US-10-283-975A-53	Sequence 53, App
23	88.2	17.6	1000	18	US-10-684-422-131	Sequence 131, App
24	88.2	17.6	1000	18	US-10-357-930-25074	Sequence 25074, A
25	88.2	17.6	1032	17	US-10-264-049-660	Sequence 6057, App
26	88.2	17.6	1385	18	US-10-723-860-6057	Sequence 6057, App
27	88.2	17.6	1385	18	US-10-057-475B-10565	Sequence 10565, A
28	85.6	17.1	350	17	US-10-154-884B-10565	Sequence 833, App
29	84.8	17.1	350	17	US-09-925-300-833	Sequence 33649, A
30	83.4	16.7	433	10	US-09-918-995-33649	Sequence 6483, App
31	82.8	16.6	383	9	US-09-960-352-6483	Sequence 19842, A
32	80.8	16.2	277	9	US-09-925-300-26	Sequence 26, App
33	79	15.8	1859	15	US-10-050-882-21	Sequence 21, App
34	78.6	15.7	978	17	US-10-115-831-70	Sequence 70, App
35	77.4	15.5	315	9	US-09-789-919-38	Sequence 38, App
36	73.4	14.7	15652	17	US-10-235-463-60	Sequence 60, App
37	73.4	14.7	134282	17	US-10-240-425-1102	Sequence 1102, App
38	71.2	14.2	490	17	US-10-242-535A-32522	Sequence 32522, A
39	71.2	14.2	490	17	US-10-085-783A-32522	Sequence 13893, A
40	69.8	14.0	413	9	US-09-960-352-13893	Sequence 170, App
41	67.8	13.6	759	9	US-09-910-943-170	Sequence 615, App
42	67.8	13.6	761	9	US-09-910-943-615	Sequence 161, App
43	67.8	13.6	760	9	US-09-910-943-161	Sequence 161, App
44	66.8	13.4	1733	18	US-10-684-422-161	Sequence 161, App
45						

ALIGNMENTS

RESULT 1
US-10-242-535A-28548
Sequence 28548, Application US/10242535A
Publication No. US20040013663A1
GENERAL INFORMATION:
APPLICANT: ChondroGene Inc.
TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
FILE REFERENCE: 4231/2005
CURRENT APPLICATION NUMBER: US/10/242,535A
CURRENT FILING DATE: 2002-09-12
PRIOR APPLICATION NUMBER: US 10/085,783
PRIOR FILING DATE: 2002-02-28
PRIOR APPLICATION NUMBER: US 60/305,340
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US 60/275,017
PRIOR FILING DATE: 2001-03-12
PRIOR APPLICATION NUMBER: US 60/271,955
NUMBER OF SEQ ID NOS: 58994
SOFTWARE: PatentIn version 3.2
SEQ ID NO 28548
LENGTH: 322
TYPE: DNA
ORGANISM: Human
FEATURES:
NAME/KEY: misc feature
LOCATION: (21)..(21)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (177)..(177)
OTHER INFORMATION: n is a, c, g, or t
US-10-242-535A-28548

```
Query Match      17.8%; Score 88.8; DB 17; Length 322;
Best Local Similarity 72.3%; Pred. No. 3.2e-19;
Matches 128; Conservative 0; Mismatches 48; Indels 1; Gaps 1;

QY      221 TTAGTTGAATGCTGTGAGTAGCGGTCTGTGTGGACAGTAAATCGAGGCGCATGT 280
DB      89 TTAATTTGAAGGTGGCGGGCGAGATGTTCTGTGTGGACAGTAAATCGAGGCGCATAC 148

QY      281 GCCGTTTAACTGATGATGAAG-CTACCCCAACGACAGAGTTTCTCAGTCAGGCCAAT 339
DB      149 ACCACTTAAGTAAACTAATGAAAGCCATATGGAACACAGGAGTTGTCAATGAGGCGAGAT 208

QY      340 CCGTTTCAGATTGATGGGCGAGCCGACAGTGAACAGACACACCTACTCAGTTGGA 396
DB      209 CAGATTCGATTTGACGGGCAACCAATCAATGAACAGACACCTGCAAGTTGGA 265

RESULT 2
US-10-085-783A-28548
; Sequence 28548, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783A
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 28548
; LENGTH: 322
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (21)..(21)
; OTHER INFORMATION: n is a, c, g, or t
; NAME/KEY: misc_feature
; LOCATION: (177)..(177)
; OTHER INFORMATION: n is a, c, g, or t
US-10-085-783A-28548

Query Match      17.8%; Score 88.8; DB 17; Length 322;
Best Local Similarity 72.3%; Pred. No. 3.2e-19;
Matches 128; Conservative 0; Mismatches 48; Indels 1; Gaps 1;

QY      221 TTAGTTGAATGCTGTGAGTAGCGGTCTGTGTGGACAGTAAATCGAGGCGCATGT 280
DB      89 TTAATTTGAAGGTGGCGGGCGAGATGTTCTGTGTGGACAGTAAATCGAGGCGCATAC 148

QY      281 GCCGTTTAACTGATGATGAAG-CTACCCCAACGACAGAGTTTCTCAGTCAGGCCAAT 339
DB      149 ACCACTTAAGTAAACTAATGAAAGCCATATGGAACAGACAGGAGTTGTCAATGAGGCGAGAT 208

QY      340 CCGTTTCAGATTGATGGGCGAGCCGACAGTGAACAGACACACCTACTCAGTTGGA 396
DB      209 CAGATTCGATTTGACGGGCAACCAATCAATGAACAGACACCTGCAAGTTGGA 265

RESULT 3
US-10-723-860-1662
; Sequence 1662, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zienlik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; FILE REFERENCE: 05882, 0193, INUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1662
; LENGTH: 557
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (228)..(228)
; OTHER INFORMATION: n is a, c, g, or t
US-10-723-860-1662

Query Match      17.8%; Score 88.8; DB 18; Length 557;
Best Local Similarity 72.3%; Pred. No. 4.1e-19;
Matches 128; Conservative 0; Mismatches 48; Indels 1; Gaps 1;

QY      221 TTAGTTGAATGCTGTGAGTAGCGGTCTGTGTGGACAGTAAATCGAGGCGCATGT 280
DB      211 TTAATTTGAAGGTGGCGGGCGAGATGTTCTGTGTGGACAGTAAATCGAGGCGCATAC 270

QY      281 GCCGTTTAACTGATGATGAAG-CTACCCCAACGACAGAGTTTCTCAGTCAGGCCAAT 339
DB      271 ACCACTTAAGTAAACTAATGAAAGCCATATGGAACAGAGGAGTTGTCAATGAGGCGAGAT 330

QY      340 CCGTTTCAGATTGATGGGCGAGCCGACAGTGAACAGACACACCTACTCAGTTGGA 396
DB      331 CAGATTCGATTTGACGGGCAACCAATCAATGAACAGACACACCTGCAAGTTGGA 387

RESULT 4
US-10-116-712-241/c
; Sequence 241, Application US/10116712
; Publication No. US20030194764A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Switzer, Ann
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.568
; CURRENT APPLICATION NUMBER: US/10/116,712
; CURRENT FILING DATE: 2002-04-07
; NUMBER OF SEQ ID NOS: 670
; SOFTWARE: FastSeq For Windows Version 4.0
; SEQ ID NO 241
; LENGTH: 300
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-116-712-241

Query Match      17.6%; Score 88.2; DB 16; Length 300;
Best Local Similarity 72.3%; Pred. No. 4.9e-19;
Matches 128; Conservative 0; Mismatches 48; Indels 1; Gaps 1;

QY      221 TTAGTTGAATGCTGTGAGTAGCGGTCTGTGTGGACAGTAAATCGAGGCGCATGT 280
DB      274 TTAATTTGAAGGTGGCGGGCGAGATGTTCTGTGTGGACAGTAAATCGAGGCGCATAC 215

QY      281 GCCGTTTAACTGATGATGAAG-CTACCCCAACGACAGAGTTTCTCAGTCAGGCCAAT 339
DB      214 ACCACTTAAGTAAACTAATGAAAGCCATATGGAACAGAGGAGTTGTCAATGAGGCGAGAT 155

QY      340 CCGTTTCAGATTGATGGGCGAGCCGACAGTGAACAGACACACCTACTCAGTTGGA 396
```


Db 214 ACCACTTAGTAAACTATGAAAGCTATTGTGAACAGAGGATTTCAATGAGGCAGAT 155

Oy 340 CCGTTTCGAGTTGATGGGACGCCGACAGTGAACAGACAACCTACTCAGTTGGA 396

Db 154 CAGATTCGAGTTGACGGGCACCAATCATATAAACAGACACACTCTGACAGTTGGA 98

RESULT 8
US-10-15

! Sequence 10573, Application US/10154884E
! Publication No. US20040005561A1

APPLICANT: Gaiger, Alexander
 APPLICANT: Algate, Paul A.
 APPLICANT: Mannion, Jane
 APPLICANT: Retter, Marc W.
 APPLICANT: Corixa Corporation
 TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
 TITLE OF INVENTION: Hematological Malignancies
 FILE REFERENCE: 014058-013521US
 CURRENT APPLICATION NUMBER: US/10/154,884B

Query Match	17.6%	Score 88.2	DB 17	Length 339
Best Local Similarity	72.3%	Pred. No. 5.2e-19		
Matches 128; Conservative		0; Mismatches 48;	Indels 1;	Gaps 1

RESULT 9
US-09-920-300A-956
; Sequence 956, Application US/09920300A
; Patent No. US20020136728A1

```

: GENERAL INFORMATION:
: APPLICANT: King, Gordon B.
: APPLICANT: Meagher, Madeleine Joy
: APPLICANT: Xu, Jiangchun
: APPLICANT: Secrist, Heather
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
: TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
: FILE REFERENCE: 210121.547
: CURRENT APPLICATION NUMBER: US/09/920,300A
: CURRENT FILING DATE: 2001-07-31
: NUMBER OF SEQ ID NOS: 1789
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 956
: LENGTH: 350
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-09-920-300A-956

```

Query Match	17.6%;	Score 88.2;	DB 9;	Length 350;
Best Local Similarity	72.3%;	Pred. No. 5.3e-19;		
Matches 128;	Conservative 0;	Mismatches 48;	Indels 1;	Gaps 1

RESULT 10
US-10-033

; Sequence 956, Application US/10033528
; Publication No. US20020131971A1

1 APPLICANT: King, Gordon E.
2 APPLICANT: Madgher, Madeline Joy
3 APPLICANT: Xu, Jiangchun
4 APPLICANT: Secrist, Heather
5 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
6 OF INFLAMMATORY COLON CANCER
7 FILE REFERENCE: 201021.547C1
8 CURRENT APPLICATION NUMBER: US/10/033,528
9 CURRENT FILING DATE: 2001-12-26
10 NUMBER OF SEQ ID NOS: 1896
11 SOFTWARE: FastSeq for Windows Version 4.0

Query Match	17.6%;	Score 88.2;	DB 13;	length 350;
Best Local Similarity	72.3%;	Pred. No. 5.3e-19;		
Matches 128;	Conservative	0;	Mismatches 48;	Indels 1;
			Gaps	1

RESULT 11

US-10-099-926-956
 ; Sequence 956, Application US/10099926
 ; Publication No. US2003016064A1
 ; GENERAL INFORMATION:
 ; APPLICANT: King, Gordon E.
 ; APPLICANT: Meagher, Madeleine Joy
 ; APPLICANT: Xu, Jiangchun
 ; APPLICANT: Secrist, Heather
 ; APPLICANT: Jiang, Yugu
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
 ; FILE REFERENCE: 210121.547C2
 ; CURRENT APPLICATION NUMBER: US/10/099,926
 ; PRIORITY FILING DATE: 2002-03-17
 ; NUMBER OF SEQ ID NOS: 1982
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 956
 ; LENGTH: 350
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-099-926-956

Query Match 17.6%; Score 88.2; DB 16; Length 350;
 Best Local Similarity 72.3%; Pred. No. 5.3e-19;
 Matches 128; Conservative 0; Mismatches 48; Indels 1; Gaps 1;

QY 221 TTAGTGAATGAGTGTGAGTGTGCTGTGTGAGACAGTAAATCGAGGCGATGT 280
 DB 77 TAATTTGAAGGTGCGGGGAGAGTGTCTGTGTGCACTTTAAAGTAAAGGCGATAC 136
 QY 281 GCCGTTTGAAGTGAAGTGAAG-CTACCCCAACAGACAGAGATTGTCACTCAAGCCCAAT 339
 DB 137 ACCACTTAGTAATCTAATGAAGCTTATTTGAACAGACAGAGATTGTCAATGAGGCGAT 196
 QY 340 CCGTTTCGAGTTGATGGGACCGGACAGTGAACAGACACCTTCACTCAATTGGA 396
 DB 197 CAGATTCCGATTGACCGGCAACCAATCAATGAACAGACACCTGCAAGATTGGA 253

RESULT 12
 US-10-242-535A-41493
 ; Sequence 41493, Application US/10242535A
 ; Publication No. US20040013663A1
 ; GENERAL INFORMATION:
 ; APPLICANT: ChondroGene Inc.
 ; APPLICANT: Iiiew, C.C.
 ; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
 ; FILE REFERENCE: 4231/2005
 ; CURRENT APPLICATION NUMBER: US/10/242,535A
 ; PRIORITY FILING DATE: 2002-09-12
 ; PRIOR APPLICATION NUMBER: US 10/085,783
 ; PRIOR FILING DATE: 2002-02-28
 ; PRIOR APPLICATION NUMBER: US 60/305,340
 ; PRIOR FILING DATE: 2001-07-13
 ; PRIOR APPLICATION NUMBER: US 60/275,017
 ; PRIOR FILING DATE: 2001-03-12
 ; PRIOR APPLICATION NUMBER: US 60/271,955
 ; PRIOR FILING DATE: 2001-02-28
 ; NUMBER OF SEQ ID NOS: 58994
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 41493
 ; LENGTH: 461
 ; TYPE: DNA
 ; ORGANISM: Human
 US-10-242-535A-41493

Query Match 17.6%; Score 88.2; DB 17; Length 461;
 Best Local Similarity 72.3%; Pred. No. 6e-19;
 Matches 128; Conservative 0; Mismatches 48; Indels 1; Gaps 1;

QY 221 TTAGTGAATGAGTGTGAGTGTGCTGTGTGAGACAGTAAATCGAGGCGATGT 280

DB 142 TAATTTGAAGGTGCGGGGAGAGTGTCTGTGTGAGACAGTAAATCGAGGCGATAC 201

QY 281 GCCGTTTGAAGTGAAGTGAAG-CTACCCCAACAGACAGAGATTGTCACTCAAGCCCAAT 339
 DB 202 ACCACTTAGTAATCTAATGAAGCTTATTTGAACAGACAGAGATTGTCAATGAGGCGAT 261

QY 340 CCGTTTCGAGTTGATGGGACCGGACAGTGAACAGACACCTTCACTCAATTGGA 396
 DB 262 CAGATTCCGATTGACCGGCAACCAATCAATGAACAGACACCTGCAAGATTGGA 318

RESULT 13

US-10-085-783A-41493
 ; Sequence 41493, Application US/10085783A
 ; Publication No. US2004003784A1
 ; GENERAL INFORMATION:
 ; APPLICANT: ChondroGene Inc.
 ; APPLICANT: Iiiew, C.C.
 ; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
 ; FILE REFERENCE: 4231/2002
 ; CURRENT APPLICATION NUMBER: US/10/085,783A
 ; PRIORITY FILING DATE: 2002-02-28
 ; PRIOR APPLICATION NUMBER: US 60/305,340
 ; PRIOR FILING DATE: 2001-07-13
 ; PRIOR APPLICATION NUMBER: US 60/275,017
 ; PRIOR FILING DATE: 2001-03-12
 ; PRIOR APPLICATION NUMBER: US 60/271,955
 ; PRIOR FILING DATE: 2001-02-28
 ; NUMBER OF SEQ ID NOS: 58994
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 41493
 ; LENGTH: 461
 ; TYPE: DNA
 ; ORGANISM: Human
 US-10-085-783A-41493

Query Match 17.6%; Score 88.2; DB 17; Length 461;
 Best Local Similarity 72.3%; Pred. No. 6e-19;
 Matches 128; Conservative 0; Mismatches 48; Indels 1; Gaps 1;

QY 221 TTAGTGAATGAGTGTGAGTGTGCTGTGTGAGACAGTAAATCGAGGCGATGT 280
 DB 142 TAATTTGAAGGTGCGGGGAGAGTGTCTGTGTGCACTTTAAAGTAAAGGCGATAC 201
 QY 281 GCCGTTTGAAGTGAAGTGAAG-CTACCCCAACAGACAGAGATTGTCACTCAAGCCCAAT 339
 DB 202 ACCACTTAGTAATCTAATGAAGCTTATTTGAACAGACAGAGATTGTCAATGAGGCGAT 261
 QY 340 CCGTTTCGAGTTGATGGGACCGGACAGTGAACAGACACCTTCACTCAATTGGA 396
 DB 262 CAGATTCCGATTGACCGGCAACCAATCAATGAACAGACACCTGCAAGATTGGA 318

RESULT 14

US-10-242-535A-39663
 ; Sequence 39663, Application US/10242535A
 ; Publication No. US20040013663A1
 ; GENERAL INFORMATION:
 ; APPLICANT: ChondroGene Inc.
 ; APPLICANT: Iiiew, C.C.
 ; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
 ; FILE REFERENCE: 4231/2005
 ; CURRENT APPLICATION NUMBER: US/10/242,535A
 ; PRIORITY FILING DATE: 2002-09-12
 ; PRIOR APPLICATION NUMBER: US 10/085,783
 ; PRIOR FILING DATE: 2002-02-28
 ; PRIOR APPLICATION NUMBER: US 60/305,340
 ; PRIOR FILING DATE: 2001-07-13
 ; PRIOR APPLICATION NUMBER: US 60/275,017
 ; PRIOR FILING DATE: 2001-03-12
 ; PRIOR APPLICATION NUMBER: US 60/271,955
 ; PRIOR FILING DATE: 2001-02-28

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 22:02:33 ; Search time 2167.44 Seconds
(without alignments)
8780.927 Million cell updates/sec

Title: US-09-926-375b-7_COPY_1_500

Perfect score: 500
Sequence: 1 tcgagagatctcttgcagc.....acaagctgccaagaccaa 500

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST.*
1: gb_est1.*
2: gb_est2.*
3: gb_hic.*
4: gb_est3.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_est1.*
9: gb_est2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	271.4	54.3	673	8 A2442128	A2442128 IM0234N09
2	104.8	21.0	902	2 BE668687	BE668687 602123353
3	101	20.2	837	2 BE208342	BE208342 601871604
4	100.2	20.0	696	2 BE208122	BE208122 601862259
5	98	19.6	797	2 BE675340	BE675340 602084513
6	98	19.6	853	2 BE698609	BE698609 602126130
7	97	19.4	710	4 BG614849	BG614849 602642387
8	97	19.4	837	2 BF028659	BF028659 601763888
9	97	19.4	866	2 BE958125	BE958125 601644972
10	96.4	19.3	744	4 B1248345	B1248345 602966952
11	96	19.2	402	2 BE568319	BE568319 601341882
12	96	19.2	605	2 BF028876	BF028876 601765249
13	96	19.2	671	2 BF184845	BF184845 601843712
14	96	19.2	680	4 BF969520	BF969520 602271842
15	96	19.2	724	4 BG114143	BG114143 602285560
16	96	19.2	738	2 BE564680	BE564680 601343236
17	96	19.2	756	2 BE568240	BE568240 601341787
18	96	19.2	763	4 BG108259	BG108259 602280245
19	96	19.2	806	2 BF207530	BF207530 601862001
20	96	19.2	829	2 BF239978	BF239978 601905182
21	96	19.2	832	2 BF968641	BF968641 602125378
22	96	19.2	844	2 BF668313	BF668313 602122518
23	96	19.2	851	2 BE564620	BE564620 601342867
24	96	19.2	858	2 BF028632	BF028632 601763858

25	96	19.2	869	2 BF790681	BF790681 602250515
26	96	19.2	874	2 BF698481	BF698481 602130669
27	96	19.2	880	2 BF207800	BF207800 601862502
28	96	19.2	887	4 BG165717	BG165717 602344554
29	96	19.2	903	2 BE866162	BE866162 601679080
30	95.4	19.1	640	4 BM353144	BM353144 1876b11.x
31	95.4	19.1	663	2 BF212445	BF212445 601813213
32	95.4	19.1	811	2 BF669919	BF669919 602118423
33	95.4	19.1	855	2 BF697745	BF697745 602131295
34	95.4	19.1	892	2 BF695449	BF695449 601851841
35	95.4	19.1	919	4 BM461727	BM461727 AGENCOURT
36	94.8	19.0	766	2 BE865640	BE865640 601677993
37	94.8	19.0	795	4 BG615729	BG615729 602643106
38	94.8	19.0	825	2 BF698120	BF698120 602130438
39	94.4	18.9	574	2 BF207801	BF207801 601862503
40	94.4	18.9	696	2 BF132136	BF132136 601821166
41	94.4	18.9	788	2 BE568246	BE568246 601341795
42	94.4	18.9	837	2 BE568054	BE568054 601341750
43	94	18.8	693	2 BE865853	BE865853 601678256
44	94	18.8	901	4 BG530611	BG530611 602560006
45	93.8	18.8	732	2 BE958229	BE958229 601644694

ALIGNMENTS

RESULT 1
A2442128/c 673 bp DNA linear GSS 03-OCT-2000
LOCUS IM0234N09R Mouse 10kb plasmid UGCLIM library Mus musculus genomic
DEFINITION clone UGCLIM0234N09 R, genomic survey sequence.

ACCESSION A2442128
VERSION A2442128
KEYWORDS GT-10566141
SOURCE GSS.
ORGANISM Mus musculus (house mouse)

REFERENCE
AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Rellly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert length: 10000 Std Error: 0.00
plate: 0234 row: N column: 09
Seq primer: CACACGGAACACGCTATGACC
Class: plasmid ends
High quality sequence stop: 673.
Location/Qualifiers

FEATURES

1. 673
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UGCLIM0234N09"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"
/clone_lib="Mouse 10kb plasmid UGCLIM library"
/note="Vector: pWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse Resource
(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (gi|473214|gb|AF12907.1), a copy-number inducible derivative of plasmid RI. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN

Query Match 54.3%; Score 271.4; DB 8; Length 673;
Best Local Similarity 99.6%; Pred. No. 5.7e-72;
Matches 272; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCGAGAGTATCTTTGTCACTGTGCTCCAAAGGGGTACTGTGCCCATAGAAAG 60
DB 273 TCGAGAGTATCTTTGTCACTGTGCTCCAAAGGGGTACTGTGCCCATAGAAAG 214
QY 61 ATCTAACTAATTAATTAATCCCTCAACGCAAAATCTTCACTAGTATAGCAGAT 120
DB 213 ATCTAACTAATTAATTAATCCCTCAACGCAAAATCTTCACTAGTATAGCAGAT 154
QY 121 TGTGAAACAAGTTCTCCAAAGAGATACAGATGATGCTGTATAGGCTGACCTGGCTG 180
DB 153 TGTGAAACAAGTTCTCCAAAGAGATACAGATGATGCTGTATAGGCTGACCTGGCTG 94
QY 181 CTGAGAGACACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
DB 93 CTGAGAGACACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 34
QY 241 AGGGTGTCTGTGGGACAGTAAATGACAGA 273
DB 33 AGGGTGTCTGTGGGACAGTAAATGACAGA 1

RESULT 2
LOCUS BF668687 902 bp mRNA linear EST 21-DEC-2000
DEFINITION 602123353F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4280532 5',
mRNA sequence.
ACCESSION BF668687
VERSION BF668687
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE
AUTHORS Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Euthera; Primates; Catarrhini; Homindae; Homo.
JOURNAL NIH-MGC http://mgi.nci.nih.gov/.
COMMENT Unpublished (1999)
National Institutes of Health, Mammalian Gene Collection (MGC)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: CLONTECH Laboratories, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM1107 row: 1 column: 13
High quality sequence stop: 597.
Location/Qualifiers

FEATURES

1..902
/organism="Homo sapiens"
/mol_type="mRNA"

ORIGIN

Query Match 21.0%; Score 104.8; DB 2; Length 902;
Best Local Similarity 70.9%; Pred. No. 1.1e-20;
Matches 139; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 201 GACTAAGAGAGCCACGGTGTATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 260
DB 142 GACTAAGAGAGCCACGGTGTATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 201
QY 261 TAGAAATCGAAGGAT 320
DB 202 TTTAAGATTAAGAGGAT 261
QY 321 GATTGTGAT 380
DB 262 GATTGTGAT 321
QY 381 CACCTACTCAGTTGGA 396
DB 322 CACCTACTCAGTTGGA 337

RESULT 3
LOCUS BF208342 837 bp mRNA linear EST 06-NOV-2000
DEFINITION 601871604F1 NIH_MGC_53 Homo sapiens cDNA clone IMAGE:4092696 5',
mRNA sequence.
ACCESSION BF208342
VERSION BF208342
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE
AUTHORS Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Euthera; Primates; Catarrhini; Homindae; Homo.
JOURNAL NIH-MGC http://mgi.nci.nih.gov/.
COMMENT Unpublished (1999)
National Institutes of Health, Mammalian Gene Collection (MGC)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: CLONTECH Laboratories, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: L1CM950 row: 9 column: 01
High quality sequence stop: 594.
Location/Qualifiers

FEATURES

1..837
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4092696"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (TI phage-resistant)"

adaptor sequence: 5'-CAGCGCCGATATGCGC-3' and 3' adaptor sequence: 5'-ATTGTAGAGGCGCGCGCGCATGATG-3' (30 BN-3' (where B = A, C, G or G (range 0.9-4.0 kb). 15/15 colonies insert size 1.65 kb (range 0.9-4.0 kb). This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

ORIGIN

Query Match	19.6%	Score 98;	DB 2;	Length 853;
Best Local Similarity	69.2%	Pred. No. 1.4e-18;		
Matches 148;	Conservative	0;	Mismatches 65;	Indels 1;
			Gaps	1

QY	201	GACTAGGAAGACCCACGGGTAGTGGTAATGATGTGGAGTGAAGGTGTTCTGTGGGACAG	26
Db	146	GACTGGAAACACGATCATATTATTTTAAAGGTGGCGGCGCAGATGGTTCTGTGGTGCAG	205
QY	261	TAGAAATCGAAGGCATGTGCGCTTTAGTGAAGCTGATGGAAG-CTACCCCAACGACAG	319
Db	206	TTTAAATTAAGGGGATACACCACTTAGTAACTAATGAAGCCCTATTGGAACGCACAG	265
QY	320	AGATTGTCTGTCAGGCCAATCCGTTTCGAGTTTGAATGGCGACCCGGAACGTGGAGACAGAC	379
Db	266	GGAATGTCAATGGCGAGATCAGATTCGGAATTTGACGGGCACCAATCAATGAAACAGAC	325
QY	380	ACACCTACTCAGTTGGAGGAAGATGAACAAT	413
Db	326	ACACCTGCACAGTTGAAATGGAGAGATGAAGAT	359

RESULT 7

LOCUS	RG614849	710 bp	mRNA	linear	EST 18-APR-2001
DEFINITION	602642387F1 NIH_MGC_61 Homo sapiens CDNA clone IMAGE:4773332 5', mRNA sequence.				

ACCESSION	EG614849	
VERSION	EG614849.1	GI:136662200
KEYWORDS	EST.	
SOURCE	Homo sapiens	(human)
ORGANISM	Homo sapiens	

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 710)	NIH-MGC	http://mgc.ncl.nih.gov/	National Institutes of Health, Mammalian Gene Collection (MGC)	
			Unpublished (1999)	
			Contact: Robert Strausberg, Ph.D.	

DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/BLNT at:
<http://image.jnl.gov>
Plate: L1CM1644 row: n column: 21
High quality sequence stop: 640.

FEATURES
source

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4773332"
/tissue_type="embryonal carcinoma"
/lab_host="DH10B (T1 phage-resistant)"
/lab_host="NH MG-61"
/notes="Oxyan: testis; Vector: pDNR-LIB (Clontech); Site_1: SfiI (ggcgccatcgccg); Site_2: SfiI (ggcgccatcgccg); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCGCATTTAGGCG-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCCGACGCCGACGATG-3' (30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average

```


Query Match 19.4%; Score 97; DB 2; Length 866;
Best Local Similarity 71.6%; Pred. No. 2.8e-18;
Matches 141; Conservative 0; Mismatches 55; Indels 1; Gaps 1;

QY 201 GACTAAGAAAGCCACGGTGTAGTTGAATGGTGTGAGTACGGTGTCTGTGGACAG 260
DB 71 GACTAGAACCAACATCATATTAAATGAAGTGGCGGGCAGATGGTTCTGTGGTGC 130
QY 261 TGAAGAAATGAGAGGATGTGCGCTTACTGAATCATGATGGAAG-CTACCCCAACGAC 319
DB 131 TTTAAGATTAAAGGATACACCATCTAGTAAACTAATGAAGCCCTATTGTGACGAC 190
QY 320 AGATTGTGAGTGGCCATCCGTTTCAGTTTGTATGGGCGGACAGTGAAGACAGAC 379
DB 191 GGATTGTCAATGAGCAGATCCGATTTGACGGGCAACCATCATTAACACGAC 250
QY 380 ACACCTACTCAGTTGA 396
DB 251 ACACCTGCACAGTTGA 267

RESULT 10
B1248345 744 bp mRNA linear EST 17-JUL-2001
LOCUS 60296952F1 NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:5116477 5',
DEFINITION mRNA sequence.
ACCESSION B1248345
VERSION B1248345.1 GI:14794150
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE NIH-MGC cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM11283 row: h column: 14
High quality sequence stop: 742.
Location/Qualifiers
1. 744
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N-3"
/db_xref="taxon:10090"
/clone="IMAGE:5116477"
/tissue_type="tumor, biopsy sample"
/dev_stage="5 months"
/lab_host="DH10B"
/clone_lib="NCI CGAP Mam2"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"

ORIGIN
Query Match 19.3%; Score 96.4; DB 4; Length 744;
Best Local Similarity 61.4%; Pred. No. 4e-18;
Matches 172; Conservative 0; Mismatches 106; Indels 2; Gaps 1;

QY 221 TTAGTTGAATGTGTGAGTAGGGTGTCTGTGGACAGTAGAAATCGAGGCAATGT 280
DB 109 TTAATTGAAGTGGCGGACAGAGATGTTCTGTGTGACAGTTAAAGATTAAAGGCAATAC 168

QY 281 GCCGTTAGTGAATGATGAGAGCTAACCCCAACGACAGATTGTCTAGTCAGGCCAATC 340
DB 169 ACCATTGTGTAACTAATGAAGAGCTATGTGAACGGAGGGTTGTCAATGAGGCAGATC 228
QY 341 CGTTTGAGTTTGATGGCCAGCCGACAGTGAAGACACACCACTTCACTCACTTGAAGAA 400
DB 229 AGATTCCGGTTTGATGGCCAGCCAGCATCAACGAACAGACACCTCGCAGTTGGAAATG 288
QY 401 GGATAGAACAAATGGCCAGCAGGAGTTGAGAGACCCCTACAGGGCCCAAGCCCTAACCA 460
DB 289 GAGATGGAAGATACATTTGATGTG--TTCCAGCAGCAGATCGAGGTGTCTACTTAAAAAG 346
QY 461 CACACCTACCACTTCATTTGACAAAGCTGCCAAGACCA 500
DB 347 GGAGCTGCTACTTACTCCAGAAATTTGTTATGACCA 386

RESULT 11
BE568319 402 bp mRNA linear EST 15-AUG-2000
LOCUS 60134182F1 NIH_MGC_53 Homo sapiens cDNA clone IMAGE:3684044 5',
DEFINITION mRNA sequence.
ACCESSION BE568319
VERSION BE568319.1 GI:9812039
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE NIH-MGC cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Clontech Laboratories, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at: image.llnl.gov
Plate: LNC370 row: c column: 21
High quality sequence stop: 402.
Location/Qualifiers
1. 402
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3684044"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_53"
/note="Organ: bladder; Vector: pDNR-LIB (Clontech);
Site 1: SfiI (ggccgctcgagc); Site 2: SfiI
(ggccatcagcc); Double-stranded cDNA was prepared from
cell line RNA. 5' and 3' adaptors were used in cloning as
follows: 5' adaptor sequence: 5'-CACGGCCATTATGCC-3' and
3' adaptor sequence:
5'-ATTCTAGAGCCGAGGCGGCGGCACATG-dT(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.55
kb (range 0.9-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA)."

ORIGIN
Query Match 19.2%; Score 96; DB 2; Length 402;
Best Local Similarity 71.6%; Pred. No. 4.4e-18;
Matches 126; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 221 TTAGTTGAATGTGTGAGTAGGGTGTCTGTGGACAGTAGAAATCGAGGCAATGT 280

D_b 159 TAAATTGAGGTGGGGGCGAGATGCTTCTGTGTCAGATTAAAGTAAAGAGCGATAC 218

O_y 281 GCCGTTTAGTAACGTAGTAGAGTATACCCCAAAGACAGATATGTCTGTCAGGCCATTC 340

D_b 219 ACCACTTAGTAACTTAATGAAAGCTTAGTGAACGACAGGATTTGTCAATGAGCGAATC 278

O_y 341 CGTTTCAGTTGATGGGCACCCGAGACGTGACAGACAACCTACTCAATTTGGA 396

D_b 279 AGATTTCGATTGACGGGCAACCATCATGAAACGACACCTGCAACAGTTGGA 334

RESULT	12
BF028876	
LOCUS	605 bp mRNA linear EST 10-OCT-2000
DEFINITION	BF028876 601765249P1 NIH_MGC_53 Homo sapiens cDNA clone IMAGE:3997332 5' , mRNA sequence.
ACCESSION	BF028876
VERSION	BF028876.1 GI:10736588
KEYWORDS	EST .
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens

REFERENCE	1 (bases 1 to 605)
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strusberg, Ph.D.

CNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 DNA distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at: image.llnl.gov
 Plate: L10CM851 row: a column: 13
 High quality sequence stop: 510.

FEATURES	Location/Qualifiers
SOURCE	1. .605

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3997332"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (TI phage-resistant)"
/clone_11b="NH MGC 53"
/notes="Organ: bladder; Vector: pDNR-LIB (Clontech);
Site 1: Sfil (ggcgccctcgccg); Site 2: Sfil
(ggcgcattcgcc); Double-stranded cDNA was prepared from
cell line RNA. 5' and 3' adaptors were used in cloning as
follows: 5' adaptor sequence: 5'-CACGCGCCATTGACC-3' and
3' adaptor sequence:
5'-ATTCTAGAGCCGAGCGGCCGACCATG-dT(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.5
kb (range 0.9-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA)."

```

ORIGIN			
Query_Match	19.2%	Score 96; DB 2;	Length 605;
Best Local Similarity	71.68%	Pred. No. 55-13;	
Matches 126; Conservative	0;	Mismatches 50;	Indels 0; Gaps 0

221 TTACTGCATGCTGTGAGTAGGGTCTCTTGCGACAGTAGAAATCGAGAGGCATCT 280

Db 90 TAAATTGAAGTGGCGGCGAGATGTTCTGTGTGAGTTTAAGATTAAGAGGCATAC 149

281 GCCGTTAGTGAAGCTGATGAGGAGCTACCCCAAAGACAGAGATTGTCAGTCAAGCCAAATC 340

Db 150 ACCACTTAGTAACTAATGAAGCCTATGTGAACGACAGGGATTGTCAATGAGGGCAGATC 209

341 CGTTTCGAGTTTGATGGCGACCGGACAGTGAACAGACACACCTTACTCAGTTTGA 396

Db 210 AGATTCCGATTGTGACGGGCAACCAATTAATGAACACACACACTCTGCACAGTTGGA 265

RESULT 13	BF184845	LOCUS	DEFINITION	ACCESION
	BF184845	671 bp	mRNA	linear
	601843712F1	NIH MGC 54	Homo sapiens	CDNA clone IMAGE:4064306 5',
	mRNA sequence,			
	BF184845			

ACCESSION	BF184845.1	GI:11063266
VERSION	BF184845.1	
KEYWORDS	EST.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 671)
NIN-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Roberts@mail.nih.gov

Email: cgap@ncrna.mcgill.ca
 Tissue Procurement: ATCC
 DNA Library Preparation: CLONETECH Laboratories, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LIML at: <http://image.llnl.gov>
 Plate: L16M898 row: h column: 03
 High quality sequence: stop: 612.

```
FEATURES
SOURCE
LOCATION/Qualifiers
1..671
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4064306"
/tissue_type="chronic myelogenous leukemia"
/lab_host="DHIO8 (T1 phage-resistant)"
```

/clone_id="nh_msc_34"
 /note="Organ: bone marrow; Vector: pDNR-LIB (Clontech);
 Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI
 (ggcattatcgcc); Double-stranded cDNA was prepared from
 cell line RNA. 5' and 3' adaptors were used in cloning as
 follows: 5' adaptor sequence: 5'-CACGGCCATTATGCGC-3' and
 3' adaptor sequence:
 5'-ATTCTAGCGCCGAGCGGCCGACATG-dt(30)BN-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size
 1.75 kb (range 0.9-4.0 kb). 15/15 colonies contained
 inserts by PCR. This library was enriched for full-length
 clones and was constructed by Clontech Laboratories (Palo
 Alto, CA)."

ORIGIN	Query Match
19.28;	Score 96; DB 2; Length 671;

Best Local Similarity 71.6%; Pred. No. 5.2e-18;
Matches 126; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

221 TTA^{CTT}GAA^{TG}GTG^{TG}GA^{GT}AGG^{GT}GT^{TT}CTG^{GG}CA^{GT}AGAA^{AT}CGAGAGC^{AT}TGT 280

Db 195 TAATTTGAGGTGCGCGGAGAGATGTTCTGTGTGCGAGTTTAAGATTTAAGAGCATAC 254

281 GCCGTTAGTGAACGTGAGACTACCCAAACGACAGAGTTGTCACTCAGGCCAATC 340

Db 255 ACCACTTACTTAACCTATGAAAGCCTATGTGAACGACAGGATTTGCTCAATGAGGCAGATC 314

341 CGTTTCGAGTTTGATGGGACGCCGACAGTGAACACACACTTACTGATTGCA 396

Db 315 AGATTCGATTGACGGGCAACCAATCAATGAAACAGACACACTGCACAGTTGCA 370

RESULT 14
BF969520

LOCUS	BF969520	680 bp	mRNA	linear	EST 22-JAN-2000
DEFINITION	602271842F1 NIH_MGC_84 Homo sapiens cDNA clone IMAGE:4359875 5', mRNA sequence.				
ACCESSION	BF969520				
VERSION	BF969520.1	GI:12336735			
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
REFERENCE	1 (bases 1 to 680)				
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .				
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)				
JOURNAL	Unpublished (1999)				
COMMENT	Contact: Robert Strausberg, Ph.D. Email: csapbs@email.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov Plate: LHAM10000 row: c column: 12 High quality sequence stop: 679.				
FEATURES	Location/Qualifiers				
source	1..680				
	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
	/clone="IMAGE:4359875"				
	/tissue_type="adrenal cortex carcinoma, cell line"				
	/lab_host="DH10B (phage-resistant)"				
	/clone_id="NIH_MGC_84"				
	/note="Organ: adrenal gland; Vector: PCMV-SPOrte; Site: 1; Not; Site 2: Salt; Cloned unidirectionally; oligo-dt primed. Average insert size 1.229 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC library."				
ORIGIN					
Query_Match	19.2%; Score 96; DB 4; Length 680;				
Best Local Similarity	71.6%; Pred. No. 5.2e-18;				
Matches	126; Conservative 0; Mismatches 50; Indels 0; Gaps 0;				
Db	221 TTAGTTGATGCTGTGAGTAGGGGTCTTCTGTGGACAGTAAATCGAGGACATGT 280				
Qy	104 TAAATTGAAGGTGGGGGCGAGGATGCTTCTGTGTGACAGTTTAAGTTAAGAGGCATAC 163				
Db	281 GCCGTTTAGTGACTGATGGAAGCTACCCCAACACACAGAGATTGTCAGTACGCCATTC 340				
Qy	164 ACCACTTAGTAACTAATGAAAGCCCTATGTGAAACACACAGGAGTTGTCAATAGGCGATTC 223				
Db	341 CGTTTCGAGTTTGATGGGCGAGCGGACAGTGAAGACACACCTACTCAGTTGGA 396				
Qy	224 AGATTCGATTTGACGGGCGACCACTCAATGAAACAGACACCTGCACAGTTGGA 279				
Db					
RESULT 15					
LOCUS	BF114143	724 bp	mRNA	linear	EST 30-JAN-2001
DEFINITION	602285560F1 NIH_MGC_86 Homo sapiens cDNA clone IMAGE:437030 5', mRNA sequence.				
ACCESSION	BF114143				
VERSION	BF114143.1	GI:12607649			
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
REFERENCE	1 (bases 1 to 724)				
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .				
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)				

```

JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
              Email: cgabbs-remail.nih.gov
              Tissue Procurement: ATCC
              cDNA Library Preparation: Life Technologies, Inc.
              CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LNL at:
              http://image.llnl.gov
              Plate: LNL0034 row: g column: 15
              High quality sequence stop: 622.
              Location/Qualifiers
                1..724
                  /organism="Homo sapiens"
                  /mol_type="mRNA"
                  /db_xref="taxon:9606"
                  /clone="IMAGE:4373030"
                  /cissue_type="osteosarcoma, cell line"
                  /lab_host="DH10B (phage-resistant)"
                  /clone_lib="NIH MGC 86"
                  /note="Organ: bone; Vector: pCMV-SPORT6; Site 1: NotI;
                  Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
                  Average insert size 1.533 kb. Library enriched for
                  full-length clones and constructed by Life Technologies.
                  Note: this is a NIH_MGC Library."

ORIGIN
Query Match          19.2%; Score 96; DB 4; Length 724;
Best Local Similarity 71.6%; Pred. No. 5,3e-18;
Matches 126; Conservative 0; Mismatches 50; Indels 0; Gaps 0

QY      221 TTAGTTCAGTGTGTGAGTACGCGGTCTGTGTGGACAGTAGAAATCGAGAGGCATGT 280
DB      13  TAAATTGAAGGTGGCGGCGGACAGATGTTCTGTGTGACGTTTAAAGTTAAGAGGCATAC 72
QY      281 GCCGTTTGAAGAACTGATGGAAGTACCCCAAACGACAGAGATTGTCAATCAGGCCAATC 340
DB      73  ACCACTTAGTAAACTAATTAAGAACCTATGTAGACGACAGGGATTGTCAATGAGGCAGATC 132
QY      341 CGTTTGCAGTTTGATGGGCGAGCCGGAACAGTAGAGACAGACACACTTACTCAGTTTGA 396
DB      133 AGATTCCGATTTGAACGGCGAACCAATCAATGATGAACAGACACACCTGCACAGTTTGA 188

Search completed: February 10, 2005, 17:32:57
Job time : 2171.44 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd

OM nucleic - nucleic search, using bw model

Run on: February 9, 2005, 21:17:58 ; Search time 2413.78 Seconds
(without alignments)
10057.285 Million cell updates/sec

Title: US-09-926-375B-7_COPY_5000_5500

Sequence: 1 ggtcagcccatgtgtctatt.....acatlaaatgccttgcga 501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 470823 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

```
1:  gb_ba: *
2:  gb_htg: *
3:  gb_in: *
4:  gb_om: *
5:  gb_ov: *
6:  gb_pat: *
7:  gb_ph: *
8:  gb_pl: *
9:  gb_pr: *
10: gb_ro: *
11: gb_rts: *
12: gb_ey: *
13: gb_un: *
14: gb_vi: *
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	501	100.0	17732	6	AX042378	AX042378 Sequence
2	501	100.0	20623	6	AX042372	AX042372 Sequence
3	501	100.0	172535	10	AL732466	AL732466 Mouse DNM
4	501	100.0	214345	2	AC107726	AC107726 Mus muscu
5	338.6	67.6	236802	2	AC103277	AC103277 Rattus no
6	333	66.5	1557	10	MMU7310	U73190 Mus musculi
7	132	26.3	136268	10	AC068486	AC068486 Mus muscu
8	132	26.3	213482	10	AC073883	AC073883 Mus muscu
9	123.4	24.6	172592	10	AC102009	AC102009 Mus muscu
10	118.6	23.7	70715	2	AC100513	AC100513 Mus muscu
11	118.6	23.7	185079	2	AC135702	AC135702 Rattus no
12	118.6	23.7	249871	2	AC118150	AC118150 Rattus no
13	117.8	23.5	177864	10	AC074315	AC074315 Mus muscu
14	117.8	23.5	192333	2	AC123698	AC123698 Mus muscu
15	114.4	22.9	193381	10	AC107453	AC107453 Mus muscu
16	114.4	22.8	205749	10	AL671187	AL671187 Mouse DNM
17	113.8	22.7	162777	2	AC124100	AC124100 Mus muscu
18	111.8	22.3	256131	2	AC125671	AC125671 Rattus no
19	111.8	22.3	296609	2	AC103542	AC103542 Rattus no

C	20	107.6	21.5	237985	2	AC125754	Rattus no
C	21	103.6	20.7	217904	2	AC015915	Mus muscu
C	22	103.6	20.7	255199	10	AL662875	Mouse DNA
C	23	103.4	20.6	210982	2	AC128082	Rattus no
C	24	103.4	20.6	241899	2	AC125909	Rattus no
C	25	101.2	20.2	244459	2	AC105937	Rattus no
C	26	101.2	20.2	278068	2	AC123167	Rattus no
C	27	100.4	20.0	223118	2	AC094074	Rattus no
C	28	99.8	19.9	237001	10	AC120124	Mus muscu
C	29	99.2	19.8	172037	2	AC119981	Mus muscu
C	30	99.2	19.8	231959	2	AC126907	Rattus no
C	31	99.2	19.8	247551	2	AC099303	Rattus no
C	32	99.2	19.8	254420	2	AC107545	Rattus no
C	33	98.8	19.7	78258	10	AL672030	Mouse DNA
C	34	98.6	19.7	210013	10	AC134791	Mus muscu
C	35	98	19.6	184501	2	AC102619	Mus muscu
C	36	97.2	19.4	216474	2	AC139414	Rattus no
C	37	97.2	19.4	240914	2	AC097252	Rattus no
C	38	97.2	19.4	260279	2	AC142066	Rattus no
C	39	95.6	19.1	257473	2	AC132663	Rattus no
C	40	95.6	19.1	297155	2	AC095921	Rattus no
C	41	95.2	19.0	221799	10	AC092856	Mus muscu
C	42	94.4	18.8	177104	10	AC091237	Mus muscu
C	43	94.4	18.8	184810	10	AC090869	Mus muscu
C	44	93.8	18.7	120734	2	AC118712	Mus muscu
C	45	93.8	18.7	144972	10	AC120122	Mus muscu

ALIGNMENTS

	RESULT 1				
LOCUS	AX042378	17732 bp	DNA	linear	PAT 23-NOV-2000
DEFINITION	Sequence 7 from Patent WO0064247.				
ACCESSION	AX042378				
VERSION	AX042378.1	GI:11340996			
KEYWORDS	.				
SOURCE	synthetic construct				
ORGANISM	synthetic construct				
	other sequences; artificial sequences.				
REFERENCE	1				
AUTHORS	Forsberg,C.W., Golovan,S. and Phillips,J.P.				
TITLE	Transgenic animals expressing salivary proteins				
JOURNAL	Patent: WO 0064247-A 7 02-NOV-2000;				
	UNIVERSITY OF GUELPH (CA)				
FEATURES	Location/Qualifiers				
source	1..17732				
	/organism="synthetic construct"				
	/mol_type="unassigned DNA"				
	/db_xref="taxon:32630"				
	/note="Lama2/APPA transgene"				
ORIGIN					
	Query Match	100.0%;	Score 501;	DB 6;	Length 17732;
	Best Local Similarity	100.0%;	Pred. No. 3.2e-128;		
	Matches 501;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Oy	1	GGTGAGCCCATGTGCTTATTTCATAGCTTTCTTTTGCTATGTTTAACAGTGGTGTCTT	60		
Db	5000	GGTAGGCCCAAGTCTTATTTCATAGCTTTCTTTTGCTATGTTTAACAGTGGTGGTGT	505		
Oy	61	GTTGGAGACCAGAGAAGCCAGTCCCGAGGCTGACAGCTGTGATACACAGGGCAGCATGA	120		
Db	5060	GTTGGAGCCAGCAGAGAAGCCAGTCCCGAGGCTGACAGCTGTGATACACAGGGCAGCATGA	511		
Oy	121	GGGTCTCAGCTTGAAGCATCTCAGGCTTGGCGAAGAAAGAACCGACACACTTCTTCA	180		
Db	5120	GGGTCTCAGCTTGAAGCATCTCAGGCTTGGCGAAGAAAGAACCGACACACTTCTTCA	517		
Oy	181	ACCAACTATGTCTTGA AAAACAACATATTATATCATATATTTGCATTATATGACAGC	240		
Db	5180	ACCAACTATGTCTTGA AAAACAACATATTATATCATATATTTGCATTATATGACAGC	523		

```

QY      241 TAAATGTACTGGGTAGATGATCTCCAGGTGGGATATCTGCAAGTGCATGAGTGGCA 300
DB      5240 TAAATGTACTGGGTAGATGATCTCCAGGTGGGATATCTCCAGGTGCATGAGTGGCA 5299
QY      301 GAGGACAGCCATGTGAGGCAAGAAAGAAATCTGGCTCAACACAGCTTAGCTCCGTGT 360
DB      5300 GAGGACAGCCATGTGAGGCAAGAAAGAAATCTGGCTCAACACAGCTTAGCTCCGTGT 5359
QY      361 GTTGGTCAAACTTTGAGAGTTTGACCAACAGCACTTATTTTGAATATTTAAACAGA 420
DB      5360 GTTGGTCAAACTTTGAGAGTTTGACCAACAGCACTTATTTTGAATATTTAAACAGA 5419
QY      421 GCACAACTTTGGGAAAAATTTTCTTATGAAAAATATTCACATTAAGCTTAAGGATGAC 480
DB      5420 GCACAACTTTGGGAAAAATTTTCTTATGAAAAATATTCACATTAAGGATGAC 5479
QY      481 TACATTAATAATGCTTTGCAA 501
DB      5480 TACATTAATAATGCTTTGCAA 5500

```

```

RESULT 2
AX042372 20623 bp DNA linear PAT 23-NOV-2000
LOCUS Sequence 1 from Patent WO0064247.
DEFINITION AX042372.1 GI:11340990
ACCESSION AX042372
VERSION AX042372.1
KEYWORDS
SOURCE
ORGANISM
other sequences; artificial sequences.

```

```

REFERENCE
1 Forsberg, C.W., Golovan, S. and Phillips, J.P.
  Transgenic animals expressing salivary proteins
  Patent: WO 0064247-A 1 02-NOV-2000;
  UNIVERSITY OF GUELPH (CA)

```

```

FEATURES
source
  1..20623
    location/Qualifiers
      /organism="synthetic construct"
      /mol_type="unassigned DNA"
      /db_xref="taxon:32630"
      /note="Lama2/Alpna plasmid"

```

ORIGIN

```

Query Match 100.0%; Score 501; DB 6; Length 20623;
Best Local Similarity 100.0%; Pred. No. 3.2e-128;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 GGTGACCCCATGTGCTTATTCATAGCTTTCTTTGCTAAGTTTACTCAGTGTGGTGT 60
DB      5000 GGTGACCCCATGTGCTTATTCATAGCTTTCTTTGCTAAGTTTACTCAGTGTGGTGT 5059
QY      61 GTTGGAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
DB      5060 GTTGGAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5119
QY      121 GGGTCTCAGCCTGAAGACAGTCAAGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
DB      5120 GGGTCTCAGCCTGAAGACAGTCAAGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAG 5179
QY      181 ACCAATATATCTTTGAAAAAACAATATATATATATATATATATATATATATATATATAT 240
DB      5180 ACCAATATATCTTTGAAAAAACAATATATATATATATATATATATATATATATATAT 5239
QY      241 TAAATGTACTCGGGTATGATGATCCAGGTGGGAGATATTCGAAGTGCATGAGTGGCA 300
DB      5240 TAAATGTACTCGGGTATGATGATCCAGGTGGGAGATATTCGAAGTGCATGAGTGGCA 5299
QY      301 GAGGACAGCCATGTGAGGCAAGAAAGAAATTCGTGCTCAACACAGCTTAGCTCCGTGT 360
DB      5300 GAGGACAGCCATGTGAGGCAAGAAAGAAATTCGTGCTCAACACAGCTTAGCTCCGTGT 5359

```

```

QY      361 GTTGGTCAAACTTTGAGAGTTTGACCAACAGCACTTATTTTGAATATTTAAACAGA 420
DB      5360 GTTGGTCAAACTTTGAGAGTTTGACCAACAGCACTTATTTTGAATATTTAAACAGA 5419
QY      421 GCACAACTTTGGGAAAAATTTTCTTATGAAAAATATTCACATTAAGCTTAAGGATGAC 480
DB      5420 GCACAACTTTGGGAAAAATTTTCTTATGAAAAATATTCACATTAAGCTTAAGGATGAC 5479
QY      481 TACATTAATAATGCTTTGCAA 501
DB      5480 TACATTAATAATGCTTTGCAA 5500

```

```

RESULT 3
AL732466 172535 bp DNA linear ROD 24-OCT-2002
LOCUS Mouse DNA sequence from clone RP23-149G14 on chromosome 2, complete
sequence.
DEFINITION AL732466.7 GI:24394996
ACCESSION AL732466
VERSION AL732466.7
KEYWORDS
SOURCE
ORGANISM
Mus musculus (house mouse)
MUS musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 172535)

```

```

REFERENCE
1 Direct Submission
  Submitted (20-OCT-2002) Wellcome Trust Sanger Institute, Hinxton,
  Cambridgeshire, CB10 1SA, UK. E-mail equities:
  humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
  On Oct 25, 2002 this sequence version replaced gi:21621699.
  ----- Genome Center
  Center: Wellcome Trust Sanger Institute
  Center code: SC
  Web site: http://www.sanger.ac.uk
  Contact: humquery@sanger.ac.uk

```

```

-----
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em., EMBL, Sw.,
SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
from the RPI-23 Mouse PAC library
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6.

```

FEATURES

```

source
  1..172535
    location/Qualifiers
      /organism="Mus musculus"
      /mol_type="genomic DNA"
      /db_xref="taxon:10090"
      /chromosome="2"
      /clone="RP23-149G14"
      /clone_lib="RPI-23"

```

ORIGIN

```

Query Match 100.0%; Score 501; DB 10; Length 172535;
Best Local Similarity 100.0%; Pred. No. 3.1e-128;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```


REFERENCE 3 (bases 1 to 236802)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On May 13, 2003 this sequence version replaced gi:25086871. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GYRO
Center clone name: CH230-235G1

----- Summary Statistics

Assembly program: Atlas 3.0;
Consensus quality: 226151 bases at least Q40
Consensus quality: 228208 bases at least Q30
Consensus quality: 228980 bases at least Q20
Estimated insert size: 234093; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

----- NOTE: Estimated insert size may differ from sequence length.
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 235405: contig of 235405 bp in length
* 235406 235505: gap of unknown length
* 235506 236802: contig of 1297 bp in length.

FEATURES

SOURCE

misc_feature

1..3075
/note="wgs_end_extension
clone_end:5p6"

misc_feature

complement(5172..5970)
/note="clone boundary
clone_end:5p6
site:ECORI

misc_feature

end_sequence:B2095899"
228453..229719
/note="wgs_contig"
233558..233880
/note="clone boundary
clone_end:77
site:ECORI

misc_feature

end_sequence:B2095897"

ORIGIN

Query Match 67.6%; Score 338.6; DB 2; Length 236802;
Best Local Similarity 86.1%; Pred. No. 4.4e-83;

Matches 433; Conservative 0; Mismatches 64; Indels 6; Gaps 5;

QY 2 GTTACGCCATGTCCTTATTCATAGCTTCTTTGCTATGTTACTCAGTGTGTTG 61
DB 175147 GTTCAGTCACTGCTTATTCATAGC-TTCTTCTGTCATTTCAGTTAGTATTTTA 175089
QY 62 TTGGACCCAGAGAGAACCCAGTCCAGGCTGACAGCTGTGATACAGGGCAGCATGAG 121
DB 175088 TTGAGACCCAGAGAACCCAGGCTGACAGCTGTGATATGATGCCAATGAG 175029
QY 122 GGTCTCAGCTTGAAGCAGTC-AGGCTGCGAGAGAGAAAGCCAGCACATTCCTTC 179
DB 175028 GGTCTTACCTTGAAGATGACAGACTGCGAGAGAAAGCCATCACATTCCTTC 174969
QY 180 AACCACTATGCTTTGAAAAAACAATATATATACATATATTCATTATGAGACAG 239
DB 174968 AACCACTGCTTTGAAAAAACAATATATATACATATATGTTATGAGACAG 174909
QY 240 CTAAATGTACTGGGTAGCATGACTCCAGTGGGATATCTGCAAGTGCATGAGTGGC 239
DB 174908 CTAAATTTACTAGGTAGCATGATATCCGTGGGATCT-TGCAAGTGCAT-AGTGGC 174851
QY 300 AGAGGACAGCCATGATGAGGAGAGAAATTCGCTTACACAGCTTACCTCTGG 359
DB 174850 AGAAGAACAGCCATGATGAGGAGAGAGACATGCTTCAACACAGCTTACCTAAGG 174791
QY 360 TGTGGTTCAAACTTTGAGAGTTGACCAACACACTTTA-TTTTGACATATTTAACA 418
DB 174790 TGTGGTTCAAACTTTGAGAGTTGACCAACACACTTTA-TTTTGATATTTAACA 174731
QY 419 GAGCACAACCTTGGAAAAAGTTTCTTATGAATATATACATTAAGCTTAAGCATG 478
DB 174730 CAGCACAAGTTTGGAAAAAGTTTCTTATGAAATATATACATTAAGCTTAAGCATG 174671
QY 479 ACTACATTAATGCTTTGCAA 501
DB 174670 ACTACATTAATGCTTTGCAA 174648

RESULT 6
LOCUS MMU73190 1557 bp DNA linear ROD 31-AUG-2001
DEFINITION Mus musculus positive acting regulatory region of the parotid
DEFINITION secretory protein (Psp) gene.
ACCESSION U73190.1 GI:1657658
VERSION U73190.1
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 1557)
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 1557)
Laursen, J., Krogh-Pedersen, H., Dagnaes-Hansen, F. and Hjorth, J. P.
The main regulatory region in the murine Psp gene is a parotid
gland enhancer
Transgenic Res. 7 (6), 413-420 (1998)

JOURNAL Transgenic Res. 7 (6), 413-420 (1998)
MEDLINE 99272895
PUBMED 10341449
REFERENCE 2 (bases 1 to 1557)
AUTHORS Laursen, J. and Poulsen, K.
TITLE Salivary gland positive acting regulatory region
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 1557)
AUTHORS Laursen, J. and Poulsen, K.
TITLE Direct Submission
JOURNAL Submitted (03-OCT-1996) Molecular and Structural Biology,
University of Aarhus, C.F. Mollers Alle, Building 130, Aarhus 8000,
Denmark

FEATURES

source

Location/Qualifiers
1..1557
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"

gene
misc_feature
/chromosome="2"
/note="located between -7.7 kb and -6.2 kb in the Mus musculus parotid secretory protein (PSP) gene"
1..>1557
/gene="PSP"
/allele="b"
/note="synonym: parotid secretory protein gene"
<1..>1557
/gene="PSP"
/allele="b"
/note="parotid secretory protein"
/function="salivary gland specific positive acting regulatory region"

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Gordon, M., Goltz, J.S. and Kucherlapati, R.
Direct Submission
Submitted (13-SEP-2000) Department of Molecular Genetics, Albert Einstein College of Medicine Genome Center, 1300 Morris Park Ave., Bronx, NY 10461, USA
4 (bases 1 to 136268)
Montgomery, K.T., Grills, G., Han, J., Lee, E., Long, J., Pomerantz, R., Ioshikhes, I.P., Shim, C., Decker, J., Thomas, E., Pereira, A., Gordon, M., Goltz, J.S. and Kucherlapati, R.
Direct Submission
Submitted (13-OCT-2000) Department of Molecular Genetics, Albert Einstein College of Medicine Genome Center, 1300 Morris Park Ave., Bronx, NY 10461, USA
On Oct 13, 2000 this sequence version replaced gi:9885874.
-----Genome Center:
Harvard Partners Genome Center
Center Code: HPGC
Web site: <http://www.hpcg.org/Sequence/mouse.html>
Contact: hpgc@mednet.harvard.edu

Query Match 66.5%; Score 333; DB 10; Length 1557;
Best Local Similarity 100.0%; Pred. No. 1.7e-81;
Matches 333; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GGTGACCCCATGTGCTTATTCATAGCTTTCTTTGCTATGTTACTCAGTGTGTTT 60
1225 GGTGACCCCATGTGCTTATTCATAGCTTTCTTTGCTATGTTACTCAGTGTGTTT 1284
61 GTTGGGACCCAGAGAGAGCCAGTCCAGGCTGACGCTGATACAGAGGAGCATGA 120
1285 GTTGGGACCCAGAGAGAGCCAGTCCAGGCTGACGCTGATACAGAGGAGCATGA 1344
121 GGGTCTCAGCTGAGAGAGCCAGTCCAGGCTGAGAGAGAGAGAGAGAGAGAGAG 180
1345 GGGTCTCAGCTGAGAGAGCCAGTCCAGGCTGAGAGAGAGAGAGAGAGAGAGAG 1404
181 ACCAATATGCTTTGAG 240
1405 ACCAATATGCTTTGAG 1464
241 TAAATATGCTTTGAG 300
1465 TAAATATGCTTTGAG 1524
301 GAGGACAGCCCATGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 333
1525 GAGGACAGCCCATGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1557

RESULT 7
AC068496/c 136268 bp DNA linear ROD 15-MAY-2002
LOCUS Mus musculus chromosome 3 clone RP23-75F13 clone RP23-75F13,
DEFINITION complete sequence.

ACCESSION AC068496
VERSION AC068496.7 GI:10800151
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE
AUTHORS
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 136268)
Montgomery, K.T., Grills, G., Han, J., Lee, E., Long, J., Pomerantz, R., Ioshikhes, I.P., Shim, C., Decker, J., Thomas, E., Pereira, A., Gordon, M., Goltz, J.S. and Kucherlapati, R.
High Throughput Mouse Sequencing

TITLE
JOURNAL
AUTHORS
Unpublished
2 (bases 1 to 136268)
Montgomery, K.T., Grills, G., Han, J., Lee, E., Long, J., Pomerantz, R., Ioshikhes, I.P., Shim, C., Decker, J., Thomas, E., Pereira, A., Gordon, M., Goltz, J.S. and Kucherlapati, R.
Direct Submission

TITLE
JOURNAL
AUTHORS
Submitted (03-MAY-2000) Department of Molecular Genetics, Albert Einstein College of Medicine Genome Center, 1300 Morris Park Ave., Bronx, NY 10461, USA
3 (bases 1 to 136268)
Montgomery, K.T., Grills, G., Han, J., Lee, E., Long, J., Pomerantz, R., Ioshikhes, I.P., Shim, C., Decker, J., Thomas, E., Pereira, A.,

REFERENCE
AUTHORS
Submitted (03-MAY-2000) Department of Molecular Genetics, Albert Einstein College of Medicine Genome Center, 1300 Morris Park Ave., Bronx, NY 10461, USA
3 (bases 1 to 136268)
Montgomery, K.T., Grills, G., Han, J., Lee, E., Long, J., Pomerantz, R., Ioshikhes, I.P., Shim, C., Decker, J., Thomas, E., Pereira, A.,

CLONE LENGTH: This sequence represents the entire insert of this clone unless otherwise noted. If there are overlapping clones, the overlaps are noted in the beginning and end of the features listing.

ANNOTATION OF FEATURES:
STS are identified using ePCR (genome Res. 7:541-550).
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished) for Human and Mouse sequences.
Genes and Regions of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to EST and cDNA sequences in Unigene. Genes demonstrate at least two exons flanked by consensus splice sites that maintain sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Attempts are made to complete double stranded sequence for all regions. All sequence is completed to a standard of coverage with a minimum of 3 reads with no ambiguities. If the sequence coverage for a region does not meet this standard, it is indicated in the annotation as low coverage. Low coverage linkages are verified by PCR product size verification or verification of forward and reverse reads from clones which span the low coverage area.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated average error rate is less than 1 per 10,000 bases using the consed quality parameters. Regions that do not meet this requirement are annotated as low quality.

-----Summary Statistics
Center project name: ACM
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 100%
Assembly program: Phrap version 0.990319
Contig length: 136268
Fraction of Phrap value < 40: 0.0011301592831773
Error Rate in Consed: 0.01 per 10,000 bases
Number of N's in consensus: 0

----- Distribution of Quality < 40 Bases:

#	Bases
1000	900
900	800
800	700
700	600
600	500
500	400
400	300
300	200
200	100
100	0

* * * * *

Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,

COMMENT	DATE	BY

All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
 Center: Whitehead Institute/MIT Center for Genome Research

Center code: MIBR

Web site: <http://www.seq.wi.mit.edu>

Contact: sequence_submissions@road.mit.edu

----- Project Information
 Center project name: L17872

Center clone name: 512_D_20

FEATURES

```

-----
Location/Qualifiers
1..172592
  /organism="Mus musculus"
  /mol_type="genomic DNA"
  /db_xref="taxon:10090"
  /chromosome="10"
  /map="10"
  /clone="RP24-512D20"
  /clone_lib="RPCI-24 Male Mouse BAC"
  complement(1..3485)
  /rpt_family="L1"
  1..4
  /note="clone boundary"
  clone_end:SP6
  site:MboI"
  complement(3473..3812)
  /rpt_family="L1"
  complement(4192..4363)
  /rpt_family="L1MA4A"
  4916..5267
  /rpt_family="RLTR10"
  complement(5278..5403)
  /rpt_family="Tigsters"
  7216..7298
  /rpt_family="MLT11"
  complement(7988..8094)
  /rpt_family="B2_Mm1"
  8095..8398
  /rpt_family="L1_MM"
  8945..8979
  /rpt_family="(TG)n"
  9231..9340
  /rpt_family="GA-rich"
  9807..10023
  /rpt_family="(TATATG)n"
  complement(10028..10311)
  /rpt_family="ORRID"
  10326..10391
  /rpt_family="(TATG)n"
  10721..10766
  /rpt_family="(CA)n"
  11239..11339
  /rpt_family="Lx8"
  complement(11162..11238)
  /rpt_family="(TCTA)n"
  complement(11340..11578)
  /rpt_family="Lx8"
  complement(12615..13469)
  /rpt_family="RLTR13B"
  complement(15004..15176)
  /rpt_family="ID_B1"
  15191..21192
  /rpt_family="L1_MM"
  21723..21742
  /rpt_family="(TTA)n"
  complement(21748..27420)
  /rpt_family="L1_MM"
  complement(27460..27567)
  /rpt_family="L1F"
  complement(27576..27696)
  /rpt_family="L1F"

```

```

repeat_region complement(27695..27812)
rpt_family="L1F"
repeat_region complement(27810..27902)
rpt_family="L1F"
repeat_region complement(27894..28079)
rpt_family="L1"
repeat_region complement(28721..30741)
rpt_family="L1_MM"
repeat_region 30746..30841
rpt_family="(TA)n"
repeat_region 30867..30899
rpt_family="(TTTA)n"
repeat_region complement(30902..30972)
rpt_family="ID4"
repeat_region 30990..31030
rpt_family="(TCTA)n"
repeat_region 31059..31127
rpt_family="(CA)n"
repeat_region 31740..32098
rpt_family="Lx8"
repeat_region 32714..33077
rpt_family="Lx8"
repeat_region 33156..33245
rpt_family="L1MA8"
33407..33996
/note="single clone coverage"
unsure
repeat_region 33518..33653
rpt_family="GA-rich"
repeat_region 33773..34036
rpt_family="L1MA8"
repeat_region 34037..34227
rpt_family="B2_Mm1"
repeat_region 34228..34263
rpt_family="L1MA8"
repeat_region complement(34281..34535)
rpt_family="B4"
repeat_region 34536..34720

```

Query Match 24.6%; Score 123.4; DB 10; Length 172592;
 Best Local Similarity 72.4%; Pred. No. 3e-23;
 Matches 160; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

```

QY 281 TGCAGTGCATGATGCGGAGAGGACAGCCATGTGAGGAGAGGAATTCGTGCTCA 340
    |||||
DB 32696 TGCAGATGATGACAGATGCGAGAGGAATTCATCATATCAAAAGGAATTCGTGTTAA 32637
    |||||
QY 341 ACACAGCTAGCTCCCTGCTGTGTGTTCAACTTTGAGCTTGACCAAGACATTTAT 400
    |||||
DB 32636 ACACGACTCGGGTTTCGTGTTGTTGAATTTCTAGAGCTGAACTCAAGACATTTAT 32577
    |||||
QY 401 TTTTGACATATTTAAACAGACACAACTTTGGAAAAAGTTTCTTATGAAAATTTATCAC 460
    |||||
DB 32576 TTTCAACATATTTAAACAGACAGTCAACTTTGAAAAAGTTCTCTGTGCAATTTGTAC 32517
    |||||
QY 461 AATTAAGCTTAAGGATGATACATTAATGCTTTGCAA 501
    |||||
DB 32516 AAGAAAGCTTAAGGATGATACACTTAACCTTTTACAA 32476
    |||||

```

RESULT 10
 AC100513
 LOCUS Mus musculus clone RP23-145F3, LOW-PASS SEQUENCE SAMPLING.
 DEFINITION AC100513
 AC100513 70715 bp DNA linear HTG 22-NOV-2001
 VERSION AC100513.1 GI:17047879
 KEYWORDS HTG; HTGS PHASED.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (Bases 1 to 70715)
 AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 TITLE Mus musculus, clone RP23-145F3

JOURNAL
REFERENCE
AUTHORS
Unpublished
2 (bases 1 to 70715)
Blitren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barta,N., Bastien,V., Boguslavsky,L., Boukhalter,B.,
Brown,A., Camarata,J., Campolano,A., Chang,J., Chazaro,B.,
Chopel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Garaya,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kellis,C., LaRocque,K.,
Lamazeas,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Margus,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrum,J.,
Menus,L., Minova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,N., Schupack,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Straus,N., Subramanian,A., Talamas,U., Testaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Triggillo,J., Vassilleu,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.U., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L15511
Center clone name: 145_F_3

* NOTE: This record contains 86 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1
* 673 772: contig of 672 bp in length
* 773 1503: contig of 731 bp in length
* 1504 1603: gap of 100 bp
* 1604 2348: contig of 745 bp in length
* 2349 2448: gap of 100 bp
* 2449 3107: contig of 659 bp in length
* 3108 3207: gap of 100 bp
* 3208 3900: contig of 693 bp in length
* 3901 4000: gap of 100 bp
* 4001 4704: contig of 704 bp in length
* 4705 4804: gap of 100 bp
* 4805 5539: contig of 735 bp in length
* 5540 5639: gap of 100 bp
* 5640 6355: contig of 716 bp in length
* 6356 6455: gap of 100 bp
* 6456 7198: contig of 743 bp in length
* 7199 7298: gap of 100 bp
* 7299 8009: contig of 711 bp in length
* 8010 8109: gap of 100 bp
* 8110 8774: contig of 665 bp in length
* 8775 8874: gap of 100 bp
* 8875 9605: contig of 731 bp in length

* 9606 9705: gap of 100 bp
* 9706 10431: contig of 726 bp in length
* 10432 10531: gap of 100 bp
* 10532 11261: contig of 730 bp in length
* 11262 11361: gap of 100 bp
* 11362 12082: contig of 721 bp in length
* 12083 12182: gap of 100 bp
* 12183 12898: contig of 716 bp in length
* 12899 12998: gap of 100 bp
* 12999 13709: contig of 711 bp in length
* 13710 13809: gap of 100 bp
* 13810 14544: contig of 735 bp in length
* 14545 14644: gap of 100 bp
* 14645 15374: contig of 730 bp in length
* 15375 15474: gap of 100 bp
* 15475 16161: contig of 742 bp in length
* 16162 16316: gap of 100 bp
* 16317 17029: contig of 713 bp in length
* 17030 17129: gap of 100 bp
* 17130 17858: contig of 729 bp in length
* 17859 17958: gap of 100 bp
* 17959 18623: contig of 665 bp in length
* 18624 18723: gap of 100 bp
* 18724 19471: contig of 748 bp in length
* 19472 19571: gap of 100 bp
* 19572 20298: contig of 727 bp in length
* 20299 20398: gap of 100 bp
* 20399 21142: contig of 744 bp in length
* 21143 21242: gap of 100 bp
* 21243 21948: contig of 706 bp in length
* 21949 22048: gap of 100 bp
* 22049 22763: contig of 715 bp in length
* 22764 22863: gap of 100 bp
* 22864 23585: contig of 722 bp in length
* 23586 23685: gap of 100 bp
* 23686 24403: contig of 718 bp in length
* 24404 24503: gap of 100 bp
* 24504 25246: contig of 743 bp in length
* 25247 25346: gap of 100 bp
* 25347 26084: contig of 738 bp in length
* 26085 26184: gap of 100 bp
* 26185 26926: contig of 742 bp in length
* 26927 27026: gap of 100 bp
* 27027 27733: contig of 707 bp in length
* 27734 27833: gap of 100 bp
* 27834 28556: contig of 723 bp in length
* 28557 28556: gap of 100 bp
* 28557 29393: contig of 737 bp in length
* 29394 29493: gap of 100 bp
* 29494 30136: contig of 643 bp in length
* 30137 30236: gap of 100 bp
* 30237 30945: contig of 708 bp in length
* 30946 31045: gap of 100 bp
* 31046 31768: contig of 723 bp in length
* 31769 31868: gap of 100 bp
* 31869 32688: contig of 820 bp in length
* 32689 32788: gap of 100 bp
* 32789 33541: contig of 753 bp in length
* 33542 33641: gap of 100 bp
* 33642 34469: contig of 728 bp in length
* 34469 34599: gap of 100 bp
* 34599 35193: contig of 724 bp in length
* 34470 35293: gap of 100 bp
* 35194 36021: contig of 728 bp in length
* 35294 36121: gap of 100 bp
* 36122 36858: contig of 737 bp in length
* 36859 36958: gap of 100 bp
* 36959 37679: contig of 721 bp in length
* 37680 37779: gap of 100 bp
* 37780 38504: contig of 725 bp in length
* 38505 38604: gap of 100 bp
* 38605 39344: contig of 740 bp in length
* 39345 39444: gap of 100 bp

```

* 39445 40175: contig of 731 bp in length
* 40176 40275: gap of 100 bp
* 40276 40935: contig of 660 bp in length
* 40936 41035: gap of 100 bp
* 41036 41749: contig of 714 bp in length
* 41750 41849: gap of 100 bp
* 41850 42582: contig of 733 bp in length
* 42583 42683: gap of 100 bp
* 42683 43418: contig of 736 bp in length
* 43419 44339: contig of 721 bp in length
* 44340 44340: gap of 100 bp
* 44340 45049: contig of 710 bp in length
* 45050 45149: gap of 100 bp
* 45150 45853: contig of 704 bp in length
* 45854 45954: gap of 100 bp
* 45954 46686: contig of 733 bp in length
* 46687 46786: gap of 100 bp
* 46787 47521: contig of 735 bp in length
* 47522 47621: gap of 100 bp
* 47622 48355: contig of 734 bp in length
* 48356 48455: gap of 100 bp
* 48456 49187: contig of 732 bp in length
* 49188 49287: gap of 100 bp
* 49288 50029: contig of 742 bp in length
* 50030 50129: gap of 100 bp
* 50130 50865: contig of 736 bp in length
* 50866 50965: gap of 100 bp
* 50966 51701: contig of 736 bp in length
* 51702 51801: gap of 100 bp
* 51802 52547: contig of 746 bp in length
* 52548 52647: gap of 100 bp
* 52648 53395: contig of 748 bp in length
* 53396 53495: gap of 100 bp
* 53496 54233: contig of 737 bp in length
* 54233 54333: gap of 100 bp
* 54333 55042: contig of 710 bp in length
* 55043 55142: gap of 100 bp
* 55143 55864: contig of 722 bp in length
* 55865 56705: contig of 741 bp in length
* 56706 56805: gap of 100 bp

```

```

Query Match      23.7%; Score 118.6; DB 2; Length 70715;
Best Local Similarity 69.9%; Pred. No. 6.5e-22;
Matches 160; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

```

```

QY 256 TAGCATATCTCCAGGTGGGATATCTGCAGTGCATGAGTGGCAGAGGACACGCCATG 315
    39012 TTGGCTTAATCTGATTTTGGGGGTCTCTCCACACACAGATGAGGAGATATGTTCTGTG 39071
QY 316 TGAAGGCAAGAAAGAAATCTGGCTCAACAGCCTAGTCCCTGGTGTGGTTCAACTTT 375
    39072 TGAAGGCAAGAAAGCAATCTGGTTAAATGTGACCTTAAGTAACTGGTACTGGTTCAACTTC 39131
Db 376 GAGAGTTGACCAACAGCACTTTATTTTGAACATATTTAAACAGAGCAACATTTGGGAA 435
    39132 TGAAGTCTGATGCTGTGAGTTATTTTAAACATATTTAAATAGACACCACTTTGGGAA 39191
QY 436 AAAGTTTCTTATGAAATTTTCACATTAAGCTTTAAGGACATGACTTACA 484
    39192 AGCATTTCTGTGATGATGATTTTACAGATTAAGCTTTAAGCCATGACTTACA 39240
Db

```

```

RESULT 11
AC135702 185079 bp DNA linear HTG 15-NOV-2002
LOCUS Rattus norvegicus clone CH230-419B1, *** SEQUENCING IN PROGRESS
DEFINITION
***, 2 unordered pieces:
AC135702
AC135702.2 GI:25007357
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE

```

ORGANISM

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE

1 (bases 1 to 185079)
Muzny, D., Marie, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alshrocks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benham, F., Bielwal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Bunay, C., Burch, P., Buttrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Diya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabriel, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, B., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogue, M., Hollins, B., Howells, S., Hulik, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kover, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewa, L., Louised, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Ming, B., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidas, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwokedi, O., Okwumodu, G., Olamunsgoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Frankoch, C., Plopper, F., Polindexter, A., Popovic, D., Primus, E., Pu, L., L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smaj, D., Sneed, A., Sodegriem, E., Song, X.-Z., Sorelle, R., Soe, J., Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabors, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhou, X., Zhou, X., von Niederhausen, A., Weis, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

TITLE

Direct Submission

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 185079)
Rat Genome Sequencing Consortium.

AUTHORS

Direct Submission

TITLE

Submitted (21-OCT-2002)

JOURNAL

Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE

3 (bases 1 to 185079)
Rat Genome Sequencing Consortium.

AUTHORS

Direct Submission

TITLE

Submitted (15-NOV-2002)

JOURNAL

Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Nov 15, 2002 this sequence version replaced gi:24182008. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas

assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: KMRB

Center clone name: CH230-419B1

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 174974 bases at least Q40

Consensus quality: 176927 bases at least Q30

Consensus quality: 178335 bases at least Q20

Estimated insert size: 182290; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

- * NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
- * NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
- * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
- * 1 85227: contig of 85227 bp in length
- * 85228 85327: gap of unknown length
- * 85328 185079: contig of 99752 bp in length.

FEATURES

source Location/Qualifiers

1..185079 "Rattus norvegicus"

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-419B1"

1..1238

/note="wgs end extension"

clone_end:Sp6

2734..3215

/note="clone boundary"

clone_end:Sp6

site:

end_sequence:BB244758"

4810..7299

/note="wgs contig"

85328..88737

/note="wgs_contig"

ORIGIN

Query Match 23.7%; Score 118.6; DB 2; Length 185079;

Best Local Similarity 72.7%; Pred. No. 6.4e-22;

Matches 168; Conservative 0; Mismatches 59; Indels 4; Gaps 1;

271 TGGGGATATCTGCAATGCGATGAGGCGAGGCGACAGCAATGAGGCGAAGAGAA 330
 65420 TGGGGATCTCCCTGCAAAATGCGACGCTGGCGAGGTAAATGATTCGTGGAGCGAAGAGAA 65479
 331 TTCTGGCTCAACACAGCTTACGCTCGTGTGTGTTCAACTTTGAGATTGACACA 390
 65480 CTCTGTTTCAAGCAAGATATAGCA-----GTGTTGATTCAACTTCAAGTCTGACCCCA 65535
 391 AGCACTTTATTTTGAATATTTAAACAGAGCAAACTTTGGAAAAAGTTTCTTATGA 450
 65536 AGCTGTTACTTTTGAATATTTAAACAGAGCAAACTTTGGAAAAAGTCTTCCACGA 65595

451 AAATATCATATAAAGCTTAAGCATGACTACATTAATAATGCTTTGCCAA 501
 65536 CATTGTCATAAATAAATTAAAGCATGACTACATGAGAACTCTTTACAA 65646

RESULT 12

AC118150/c

LOCUS

DEFINITION

AC118150

Rattus norvegicus clone CH230-107K7, WORKING DRAFT SEQUENCE, 3

unordered pieces.

AC118150

AC118150.5 GI:25073571

HTG; HTGS PHASE1; HTGS DRAFT; HTGS_PULOTOP.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 249871)

Muzny D, Marie, Metzker M, Lee, Abramson S, Adams C, Alder J, Allen C, Allen H, Albrooks S, Amin A, Angiano D, Anyalebech V, Aoyagi A, Ayodeji M, Baca E, Baden H, Baldwin D, Bandaranaike D, Barber M, Barnstead M, Benahmed F, Bismail K, Blair J, Blankenburg K, Blyth P, Brown M, Bryant N, Buhay C, Burch P, Burrell K, Calderon E, Cardenas V, Carter K, Cavazos I, Caesar H, Center A, Chacko J, Chavez D, Chen G, Chen R, Chen Y, Chen Z, Chu J, Cleveland C, Cockrell R, Cox C, Coyle M, Cree A, d'Souza L, Davila M, Davis C, Davy-Carroll J, De Ande C, Dederich D, Delgado O, Denison S, Derramo C, Ding Y, Dinh H, Divya K, Draper H, Dugan-Rocha S, Dunn A, Durbin K, Duval B, Eaves K, Egan A, Escotto M, Eugene C, Evans C, Falls T, Fan G, Fernandez S, Finley M, Flagg N, Forbes L, Foster M, Foster P, Frazer C, Gabisi A, Gante R, Garcia A, Garner T, Garza M, Gebregeorgis E, Geer K, Gill R, Grady M, Guerra W, Guevara W, Gunaratne P, Haaland W, Hamill C, Hamilton C, Hamilton K, Harvey Y, Havlak P, Hawes A, Henderson N, Hernandez J, Hernandez B, Hines S, Hladun S, Hume J, Idlebird D, Jackson A, Jackson L, Jacob L, Jiang H, Johnson B, Johnson R, Jolivet A, Karpathy S, Kelly S, Kelly S, Khan Z, King L, Kovar C, Kowalski S, Kraft C, Lebow H, Levan J, Lewis L, Li Z, Liu J, Liu J, Liu W, Liu Y, London P, Longacre S, Lopez J, Lorenshew L, Louisedge H, Lozano R, Lu X, Ma J, Maheshwari M, Mahindaratne M, Mahmood M, Malloy K, Mangum A, Mangum B, Mapua P, Martin K, Martin R, Martinez E, Mawhinney S, McLeod M, McNeill T, Meenen E, Milosavljevic A, Miner G, Minja E, Montemayor J, Moore S, Morgan M, Morris K, Morris S, Mundada M, Murphy M, Nair L, Nankervis C, Neal D, Newton N, Nguyen N, Norris S, Nocklelemb O, Okmonu G, Olarunpungoon A, Pal S, Parks K, Pasternak S, Paul H, Perez A, Perez L, Pfannkuch C, Plopper F, Poindexter A, Popovic D, Primus E, Pu L, Puazo M, Quito J, Rachlin E, Reeves K, Regier M, Reigh R, Reilly B, Reilly M, Ren Y, Reuter M, Richards S, Riggs F, Rives C, Rodkey T, Rojas A, Rose M, Rose R, Ruiz S, Sanders W, Savery G, Scherer S, Scott G, Shatman S, Shen H, Shetty J, Shvartsbeyn A, Sison I, Sitter C, D. Smaj D, Sneed J, Sodergren E, Song X, Z, Sorelle R, Soza J, Steidle M, Strong R, Sutton A, Svatek A, Taber P, Taylor C, Taylor T, Thomas N, Thomas S, Tingey A, Trejos Z, Uman K, Vais R, Ver V, Villasana D, Waldron L, Walker B, Wang J, Wang Q, Wang S, Warren J, Warren R, Wei X, White F, Williams G, Wilson R, Wlezyk R, Wooden H, Worley K, Wright D, Wright R, Wu J, Yakub S, Yen J, Yoon L, Yoon V, Yu F, Zhang J, Zhou J, Zhou S, Zhao S, Dunn D, von Niederhausern A, Weiss R, Smith D, Holt R, Smith H, O, Weinstock G, and Gibbs R.A.

TITLE

Unpublished

2 (bases 1 to 249871)

Worley K.C.

Direct Submission

REFERENCE

AUTHORS

TITLE

JOURNAL Submitted (14-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 249871)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On Nov 19, 2002 this sequence version replaced gi:23673599.
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rac/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 Project Information
 Center project name: GUNH
 Center clone name: CH230-107K7

----- Summary Statistics
 Assembly program: Phrap; version 0.990329
 Consensus quality: 213554 bases at least Q40
 Consensus quality: 216552 bases at least Q30
 Consensus quality: 218611 bases at least Q20
 Estimated insert size: 217784; sum-of-contigs estimation
 Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/cenbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 3 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 247137: contig of 247137 bp in length
 * 247138 247237: gap of unknown length
 * 247238 248565: contig of 1328 bp in length
 * 248566 248665: gap of unknown length
 * 248666 249871: contig of 1206 bp in length.

FEATURES
 source
 1. .249871
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-107K7"

misc_feature
 1. .1763
 /note="wgs_end_extension
 clone_end:T7"
 6697. .7589
 /note="clone_boundary
 clone_end:T7
 site:
 end_sequence:BH270256"

misc_feature
 /note="wgs_end_extension
 clone_end:T7"
 6697. .7589
 /note="clone_boundary
 clone_end:T7
 site:
 end_sequence:BH270256"

ORIGIN
 Query Match 23.7%; Score 118.6; DB 2; Length 249871;
 Best Local Similarity 72.7%; Pred. No. 6.4e-22;
 Matches 168; Conservative 0; Mismatches 59; Indels 4; Gaps 1;

QY 271 TGGGATATCTGCAAGTGGCATGAGTGGCAGAGGACAGCAATGTGAGGCAAGAGCA 330
 |||||
 Db 202154 TGGGGTTCCTCGCAATGCGACAGCTGAGGGTATGATTCGCGAGGCAAGAGGAA 202095
 |||||

QY 331 TTCTGGCTCAACACAGCTTAGCTCCCTGGTGTGGTTCAACTTGAGACTTGACCA 390
 |||||
 Db 202094 CTCTGTTGAGCAAGATATGACCA----GTGTGATTCAACTTAGAGCTGACCCCA 202039
 |||||

QY 391 AGACCTTATTTTGTGACATTTTAAACAGACCACTTTGGGAAAAGTTTCTATGA 450
 |||||
 Db 202038 AGCTGTTTCTTTGATATTTTAAACAGACCACTTTGAAAACGCTTTCCACGA 201979
 |||||

QY 451 AATTATCAATTAAGCTTAAGGACATGACTACATTAATGCTTTGCA 501
 |||||
 Db 201978 CAATGTGCAATTAATTAATTAAGGACATGACTACATGGAACCTCTTACA 201928
 |||||

RESULT 13
 LOCUS AC074315
 DEFINITION Mus musculus clone rp23-445c23 map 15 strain C57BL/6J, complete sequence.

ACCESSION AC074315
 VERSION AC074315.77 GI:48331781
 KEYWORDS HTG.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 177864)
 Do T., Do A., Hauser, L. and Roe, B.A.
 Mus musculus BAC clone rp23-445c23
 Unpublished
 2 (bases 1 to 177864)
 Do T., Do A., Hauser, L. and Roe, B.A.
 Direct Submission
 Submitted (26-JUL-2000) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
 3 (bases 1 to 177864)
 Do T., Do A., Hauser, L. and Roe, B.A.
 Direct Submission
 Submitted (25-MAY-2004) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
 4 (bases 1 to 177864)
 Do T., Do A., Hauser, L. and Roe, B.A.
 Direct Submission
 Submitted (25-MAY-2004) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
 5 (bases 1 to 177864)
 Do T., Do A., Hauser, L. and Roe, B.A.
 Direct Submission
 Submitted (26-MAY-2004) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
 6 (bases 1 to 177864)
 Do T., Do A., Hauser, L. and Roe, B.A.
 Direct Submission
 Submitted (28-MAY-2004) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
 7 (bases 1 to 177864)
 Do T., Do A., Hauser, L. and Roe, B.A.
 Direct Submission
 Submitted (03-JUN-2004) Department of Chemistry And Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
 8 (bases 1 to 177864)
 Do T., Do A., Hauser, L. and Roe, B.A.
 Direct Submission
 Submitted (05-JUN-2004) Department of Chemistry And Biochemistry,

COMMENT
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
On Jun 5, 2004 this sequence version replaced gi:48058106.

----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR

FEATURES

```

Location/Qualifiers
1..177864
/organism="Mus musculus"
/mol_type="genomic DNA"
/scrain="C57BL/6J"
/db_xref="taxon:10090"
/mab="15"
/cname="IP23-44Sc23"
clone_11b="R1C1 - 23 Female (C57BL/6J) Mouse BAC Library"

```

ORIGIN

Query Match	23.5%	Score 117.8;	DB 10;	Length 177864;
Best Local Similarity	68.2%	Pred. No. 1.1e-21;		
Matches 178;	Conservative	0;	Mismatches 82;	Indels 1;
			Gaps	1;

QY	241	TTAAATGTA	CTCGGGTGA	CACTGA	CTCCAGGTGGGGAA	TATCTGCAAGTGCATAGAGGCA	300
Db	46616	TTAACTTA <td>CTTTTATTA</td> <td>CTTTTGTC</td> <td>CAATGACAGGGGCT</td> <td>CTTAAGCTCACTAGTTGGCA</td> <td>46675</td>	CTTTTATTA	CTTTTGTC	CAATGACAGGGGCT	CTTAAGCTCACTAGTTGGCA	46675
QY	301	GAGGAGAC	CCCAATGTAGG	CAAGAGAA	TTTGCTCA	CAACAGTTAGCTCCTGGT	360
Db	46676	GAGAAATTA	ACCAACAAGG	GCAGAAAGAA	GTCTTGTC	GCATGACTTAAGTTAGGTAG- 46734	
QY	361	GTGTGTTCA	ACTTGA	AGTTTGA	CAACAAGCA	CTTATTTTGCATTTTAAACAG	420
Db	46735	GCTAGTTCA	ACTTCTGG	AGGTACATCA	AGCAATTTATTTT	TAGCATTTTATATACA 46794	
QY	421	GCACAACTT	GGGAAAA	AGTTTCTTA	TGAAAAATTA	CAACATPAAAGCTTAAGGCATGAC	480
Db	46795	GTACAACTT	GGAAATAC	AGTCTTCT	GCATGACAAATATCA	AAATPAAAGCTTAAGGCATGAC 46854	
QY	481	TACATTTAA	ATGCTCTTGCA	501			
Db	46855	TATGTCAAA	ATTCCTTTACAA	46875			

RESULT	14
AC123698/c	
LOCUS	AC123698
DEFINITION	Mus musculus chromosome 15 clone RP2-331P14 map 15, *** SEQUENCING IN PROGRESS ***.
	19233 bp DNA linear HTG 23-SEP-2004

ACCESSION AC123698
VERSION AC123698.1
KEYWORDS HTG: HTGS PHASE2; HTGS_FULTOP; HTGS_ACTIVEFIN
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE AUTHORS TITLE JOURNAL REFERENCE AUTHORS	
1 (bases 1 to 192333)	Birren, B., Nusbaum, C. and Lander, E.
Mus musculus chromosome 15, clone RP23-331P1	
Unpublished	
2 (bases 1 to 192333)	Birren, B., Linton, L., Nusbaum, C., Lander, E.,

TITLE Direct Submersion
JOURNAL Submitted (01-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (pages 1 to 192333)
AUTHORS Birtn, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., ... et al. (1999) Nature 398: 273-279

TITTLE
JOURNAL

JOURNAL Submitted (23-SEP-2004) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Sep 23, 2004 this sequence version replaced gi:52077830.

Center: Whitehead Institute/MIT Center for Genome Research
 Center code: MMRB
 Web site: <http://www.sag.wi.mit.edu>
 Contact: sequence_submissions@broad.mit.edu

 Project Information
 Center project name: L66429
 Center clone name: 331_P_14

```
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
  1 192333: contig of 192333 bp in length.
```

FEATURES

```

1..192333
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="15"
/map="15"
/clone="RP23-331P14"
/clone_1b="BPC1-23 Female Mouse BAC"

```

ORIGIN

Query Match	23.58; Score 117.8; DB 2; Length 1923337
-------------	--

Best Local Similarity 68.2%; Pred. No. 1.1e-21;
Matches 178; Conservative 0; Mismatches 82; Indels 1; Gaps 1;

QY 241 TAAATGTAAGTGGGAGTACGATCAGGAGTGGGATATCTGCAGATGTCAGTGGCA 300
Db 2764 TTTACTTACTTATTTATTTCTTTGTCCATTCAGGGGCTCTGTAAGCTCTACTGTTGGCA 2705
QY 301 GAGGAGCAGCCAAATGTGAGCAAGAAAGAAATTCGTGCTCAACACAGTTAGTCCCTGT 360
Db 2704 GAGAAATTAACCAAGAGGCAAGAAAGAGTGTGTTTGTGATGACTTGTAGCTAGCTAG- 2646
QY 361 GTTGTGTTCAACTTTGAGATTTGACCAAGACACTTATTTTGTGACATTTTAAACAGA 420
Db 2645 GCTAGTTCAACTTGTGAGTGTGACATCAAGCAATTTATTTTGTGATTTATATACA 2586
QY 421 GCACAACTTGGGAAAAAGTTTCTTGTGAAATTTACAAATTAAGCTTAAGGCATGAC 480
Db 2585 GTTCAACCTTGGATACAGTCTTCTCTGACAAATTAAGCTTAAGGCATGAC 2526
QY 481 TACATTAATAATGCTTTGCAA 501
Db 2525 TATGTCAAATTCCTTACAA 2505

RESULT 15
AC107453 193381 bp DNA linear ROD 30-SEP-2003
LOCUS Mus musculus chromosome 15, clone RP23-315F3, complete sequence.
AC107453
AC107453.14 GI:33504630
VERSION HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
TITLE Birren, B., Nussbaum, C. and Lander, E.
JOURNAL Mus musculus chromosome 15, clone RP23-315F3
REFERENCE
TITLE 2 (bases 1 to 193381)
AUTHORS Birren, B., Linton, L., Nussbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barina, N., Bastien, V., Boguslavsky, L., Boukhalter, B.,
Brown, A., Camarata, J., Campiano, A., Chang, J., Chazaro, B.,
Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., Lavoque, K.,
Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McSheeters, R., Meldrim, J.,
Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunhthang, P., Pierre, N., Pollard, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Stojanovic, N.,
Stearns, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Straus, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL Direct Submission
Submitted (21-JAN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 193381)

REFERENCE
AUTHORS Birren, B., Nussbaum, C., Lander, E., Abouelleil, A., Allen, N.,
Anderson, M., Arachchi, H.M., Barina, N., Bastien, V., Bloom, T.,
Boguslavsky, L., Boukhalter, B., Camarata, J., Chang, J., Choepel, Y.,
Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K.,
Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,
Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S.,
Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,

TITLE
JOURNAL Direct Submission
Submitted (27-JUL-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 193381)

REFERENCE
AUTHORS Birren, B., Nussbaum, C., Lander, E., Abouelleil, A., Allen, N.,
Anderson, M., Arachchi, H.M., Barina, N., Bastien, V., Bloom, T.,
Boguslavsky, L., Boukhalter, B., Camarata, J., Chang, J., Choepel, Y.,
Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K.,
Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,
Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S.,
Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,
Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,
Lindblad-Toh, K., Liu, X., Liu, A., Mabbitt, R., Maclean, C.,
Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,
Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J.,
Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., Peterson, K., Phunhthang, P., Pierre, N.,
Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Schauer, S., Schnupack, R., Seaman, S., Severy, P., Smith, C.,
Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,
Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M.,
Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL Direct Submission
Submitted (30-SEP-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 8, 2003 this sequence version replaced gi:33385400.
All repeats were identified using RepeatMasker:
Smit, A.P.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

COMMENT
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 20:12:02 ; Search time 341.632 Seconds
(without alignments)
8681.254 Million cell updates/sec

Title: US-09-926-375B-7_COPY_5000_5500

Perfect score: 501

Sequence: 1 ggtcagcccatcgtcttalt.....acatcaaatgccttgcac 501

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*

1: geneseq1980s:.*
2: geneseq1990s:.*
3: geneseq2000s:.*
4: geneseq2001as:.*
5: geneseq2001bs:.*
6: geneseq2002as:.*
7: geneseq2002bs:.*
8: geneseq2003as:.*
9: geneseq2003bs:.*
10: geneseq2003cs:.*
11: geneseq2003ds:.*
12: geneseq2004as:.*
13: geneseq2004bs:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	501	100.0	17732	3 AAC68300	Aac68300 Lama2/APP
2	501	100.0	20623	3 AAC68294	Aac68294 Lama2/APP
3	72	14.4	211257	11 ACN44200	Acn44200 Mouse gen
4	69.2	13.8	233060	11 ACN43912	Acn43912 Mouse gen
5	65.6	13.1	89824	13 ABD32895	Abd32895 Mouse can
6	65.2	13.0	110000	8 ABX16390	Abx16390 Mouse hig
7	63.6	12.7	48133	11 ACN44860	Acn44860 Mouse gen
8	61.6	12.3	2893	6 AB199246	Ab199246 Mouse isc
9	61.6	12.3	145068	13 ABD33090	Abd33090 Murine ca
10	61.4	12.3	67832	9 ADA02801	Ada02801 Mouse itk
11	61.4	12.3	67832	10 ADB72539	Adb72539 Mouse itk
12	61.4	12.3	67832	10 ADC85281	Adc85281 Mouse itk
13	61.4	12.3	67832	12 ADM74396	Adm74396 Murine ca
14	60.8	12.1	210528	11 ACN44040	Acn44040 Mouse gen
15	60.4	12.1	121129	13 ABD33446	Abd33446 Murine ca
16	60	12.0	110000	13 ABD32627	Abd32627 Continuation (3 of
17	58.6	11.7	96599	9 ADA02933	Ada02933 Mouse Bra
18	58.6	11.7	96599	10 ADB72671	Adb72671 Mouse Bra
19	58.6	11.7	96599	10 ADC85413	Adc85413 Mouse Bra
20	58.6	11.7	96599	12 ADM74528	Adm74528 Murine ca

ALIGNMENTS

RESULT 1	
AAc68300	standard; DNA; 17732 BP.
ID	AAc68300 standard; DNA; 17732 BP.
XX	
AC	AAc68300;
XX	
DT	15-SEP-2003 (revised)
DT	20-FEB-2001 (first entry)
XX	
DE	Lama2/APPa plasmid coding sequence.
XX	
KW	Transgenic animal; salivary protein; phytase; phosphorus; animal growth;
KW	environmental pollution; pig; ds.
XX	
OS	Mus musculus.
OS	Escherichia coli.
OS	Chimeric.
XX	
PN	WO200064247-A1.
XX	
AC	02-NOV-2000.
PD	
XX	
PF	20-APR-2000; 2000WO-CA000430.
XX	
PR	23-APR-1999; 99US-0130508P.
XX	
PA	(UYGU-) UNIV GUELPH.
XX	
PI	Forsberg CW, Golovan S, Phillips JP;
XX	
WP	2000-687245/67.
DR	P-PSDB; AAB36263.
DR	
XX	
PT	Transgenic non-human animal for gastrointestinal tract specific
PT	expression of a protein, preferably phytase, comprises a nucleic acid
PT	sequence including a heterologous transgene construct encoding the
PT	protein.
XX	
PS	Claim 14; Fig 23; 152p; English.
XX	
CC	The present invention provides transgenic animals which produce desired
CC	proteins, in this case pigs which express phytase in the salivary
CC	gland. Low phytase production levels result in phytate in the diet being

21	58	11.6	75101	11 ACN44672	Acn44672 Mouse gen
22	57	11.4	80268	13 ABD32951	Abd32951 Mouse can
23	54.4	10.9	52256	11 ACN44780	Acn44780 Mouse gen
24	53.2	10.6	55996	9 ADA02741	Ada02741 Mouse Trif
25	53.2	10.6	55996	10 ADB72479	Adb72479 Mouse Trif
26	53.2	10.6	55996	10 ADC85221	Adc85221 Mouse Trif
27	53.2	10.6	55996	12 ADM74336	Adm74336 Murine ca
28	53	10.6	204803	12 ADQ97348	Adq97348 Mouse can
29	50	10.0	247461	13 ABD33153	Abd33153 Murine ca
30	49.6	9.9	23889	13 ABD33135	Abd33135 Murine ca
31	49.6	9.9	23889	13 ABD32568	Abd32568 Mouse can
32	49.2	9.8	96289	13 ABD33205	Abd33205 Murine ca
33	47.8	9.5	52637	13 ADB67083	Adb67083 Mouse can
34	47.4	9.5	54552	11 ACN44716	Acn44716 Mouse gen
35	46.6	9.3	173810	6 ABN85752	Abn85752 Mouse chr
36	40	8.0	58053	13 ABD32663	Abd32663 Mouse can
37	39.2	7.8	2000	8 ADA71938	Ada71938 Rice gene
38	39.2	7.8	3892	12 ADO35507	Ado35507 Novel mou
39	38.6	7.7	435	5 ABV54350	Abv54350 Human pro
40	38.6	7.7	44728	13 ABD32556	Abd32556 Mouse can
41	38.6	7.7	191284	12 ADQ97957	Adq97957 Mouse gen
42	37.8	7.5	27920	11 ACN44776	Acn44776 Mouse gen
43	37.6	7.5	50781	11 ACN44360	Acn44360 Mouse gen
44	37.4	7.5	846	6 ABK90144	Abk90144 Barley ge
45	36.6	7.3	114454	12 ADQ97873	Adq97873 Mouse can

CC excreted and causing phosphorus contamination in water, as well as
CC reducing the growth of animals. The invention provides a number of
CC transgenes containing the *E. coli* *ADPA* phylyase coding sequence. (Updated
CC on 15-SEP-2003 to standardise OS field)
XX
SQ Sequence 17732 BP; 4719 A; 4125 C; 4168 G; 4719 T; 0 U; 1 Other;

Query Match	Score	DB	Length
100.0%	501	3	17732

```
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY	1	GGTACGCCATGTGGTTATTCATAGCTTTCTTTGGCATATGTTACTAGGTGAGTGT	60
Db	5000	GGTCAGCCCAATGTCTTATTCATAGCTTTCTTTGGCATATGTTACTAGGTGAGTGT	5059
QY	61	GTTGGAGCCACAGACAGACCAGTCCAGGCTGACAGCTGTGGAATACAGGGCAGCATGA	120
Db	5060	GTTGGAGCCACAGACAGACCAGTCCAGGCTGACAGCTGTGGAATACAGGGCAGCATGA	5119
QY	121	GGGTCTCAGCCTGAAGAGTCAAGGCTGGCAGAAAGAAAGCCAGCACATTCCTTCA	180
Db	5120	GGGTCTCAGCCTGAAGAGTCAAGGCTGGCAGAAAGAAAGCCAGCACATTCCTTCA	5179
QY	181	ACCAACTATGTCTTGAAAAACAAACATATTATACATATATGCAATTATATGACACAGC	240
Db	5180	ACCAACTATGTCTTGAAAAACAAACATATATATATACATATATGCAATTATATGACACAGC	5239
QY	241	TAAATGTATCTCGGGTAGCATGATCTCCAGGTGGGGATATCTGCAGATGCTCATGAGTGCAC	300
Db	5240	TAAATGTATCTCGGGTAGCATGATCTCCAGGTGGGGATATCTGCAGATGCTCATGAGTGCAC	5299
QY	301	GAGGACACAGCCCATGTGAGGCGAAGAAGAAATTCTGGCTCAACACAGCTTAACTCCTGGT	360
Db	5300	GAGGACACAGCCCATGTGAGGCGAAGAAGAAATTCTGGCTCAACACAGCTTAACTCCTGGT	5358
QY	361	GTTGGTTCAAACTTTGAGAGTTTGACCAACAACCATTTATTTTACATATTTAAACAGA	420
Db	5360	GTTGGTTCAAACTTTGAGAGTTTGACCAACAACCATTTATTTTACATATTTAAACAGA	5419
QY	421	GCACAACCTTGGGAAAAAGTTTCTTATGAAAAATATATCACAATAAAGCTTAAAGCATAC	480
Db	5420	GCACAACCTTGGGAAAAAGTTTCTTATGAAAAATATATCACAATAAAGCTTAAAGCATAC	5479
QY	481	TACATTAAAAATCCCTTTGCAA 501	
Db	5480	TACATTAAAAATCCCTTTGCAA 5500	

RESULT 2

AAC68294
ID AAC68294 standard; DNA; 20623 BP.

XX	
AC	AAC68294;
XX	
DT	15-SEP-2003 (revised)
DT	20-FEB-2001 (first entry)
XX	
DE	Lama2/APPA plasmid coding sequence.
XX	
KW	Transgenic animal; salivary protein; phytase; phosphorus; animal growth,
KW	environmental pollution; pig; ds.
XX	
OS	Mus musculus.
OS	Escherichia coli.
OS	Chimeric.
XX	
PN	WO200064247-A1.
XX	
PD	02-NOV-2000.
XX	
PF	20-APR-2000; 2000WO-CA000430.
XX	
PR	23-APR-1999; 99US-0130508P.

XX
PA (UYGU-) UNIV GUELPH.
XX
XX
PI Forsberg CW, Golovan S, Phillips JP,
XX
DR WPI; 2000-687245/67.
DR P-PsDB; AAB36257.

Transgenic non-human animal for gastrointestinal tract specific expression of a protein, preferably phytase, comprises a nucleic acid sequence including a heterologous transgene construct encoding the protein..

Claim 56; Fig 5; 152pp; English.

The present invention provides transgenic animals which produce desired proteins, in this case pigs which express phytase in the salivary gland. Low phytase production levels result in phytate in the diet being excreted and causing phosphorus contamination in water, as well as reducing the growth of animals. The invention provides a number of transgenes containing the E. coli Alpha phytase coding sequence. (Updated on 15-SEP-2003 to standardise CC field)

SQ Sequence 20623 BP; 5449 A; 4847 C; 4902 G; 5424 T; 0 U; 1 Other;

Query Match	100.0%;	Score 501;	DB 3;	Length 20623;
Best Local Similarity	100.0%;	Pred. No. 4.9e-138;		
Matches 501;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

Qy	I	GGTCAGCCCAATGCTCTTAATTCGAAAGCTTTCTTTGCAATGTTACCTCAGGTGGAGTGT	60
Db	5000	GGTCAGCCCAATGCTCTTAATTCGAAAGCTTTCTTTGCAATGTTACCTCAGGTGGAGTGT	5059
Qy	61	GTTGGGACCCAGACGAAGACGAGTCCCAAGGCTGACAGCTGTGGATACA CAGGGCAGACATGA	120
Db	5060	GTTGGGACCCAGACGAAGACGAGTCCCAAGGCTGACAGCTGTGGATACA CAGGGCAGACATGA	5119
Qy	121	GGGTCCTCAGCCTGAGACAGTCAGGCTGGCAGAAAGAAAAGACGACACACATTTCTTTCA	180
Db	5120	GGGTCCTCAGCCTGAGACAGTCAGGCTGGCAGAAAGAAAAGACGACACACATTTCTTTCA	5179
Qy	181	ACCAACCTATGTCCTTGA AAAACAAA CAAATTTATATCAATATATTGCATTTATGACACGC	240
Db	5180	ACCAACCTATGTCCTTGA AAAACAAA CAAATTTATATCAATATATTGCATTTATGACACGC	5239
Qy	241	TAAATGTACTCGGCTAGCATGACTCCACGAGGGGGAATTCGCAAGGCCATGAGTGCA	300
Db	5240	TAAATGTACTCGGCTAGCATGACTCCACGAGGGGGAATTCGCAAGGCCATGAGTGCA	5299
Qy	301	GAGGACAGCCCAATGTGAGGCAGAAGAAATTCGGCTCAACACAGCTTAGCTCCGTGT	360
Db	5300	GAGGACAGCCCAATGTGAGGCAGAAGAAATTCGGCTCAACACAGCTTAGCTCCGTGT	5359
Qy	361	GTTGGTCAAACTTTGAGAGTTTGCACCAAGCACTTATTTTTTGACATATTTAAACAGA	420
Db	5360	GTTGGTCAAACTTTGAGAGTTTGCACCAAGCACTTATTTTTTGACATATTTAAACAGA	5419
Qy	421	GCACAACTTTGGGAAAAAGTTTCTTATGAAAAATATACAAATAAGCTTAAAGCATGAC	480
Db	5420	GCACAACTTTGGGAAAAAGTTTCTTATGAAAAATATACAAATAAGCTTAAAGCATGAC	5479
Qy	481	TACATTAAAAAGCTTTGCCAA 501	
Db	5480	TACATTAAAAAGCTTTGCCAA 5500	

RESULT 3

ACN44200/c
ID ACN44200 standard; DNA; 211257 BP

XX	ACN44200;
AC	
XX	
DT	18-NOV-2004 (first entry)

XX Mouse genomic sequence mCG15870.
XX Cytostatic; carcinoma; lymphoma; cancer; murine; gene; ss.
XX Mus musculus.
XX WO2003073826-A2.
XX 12-SEP-2003.
XX 28-FEB-2003; 2003WO-US006235.
XX 01-MAR-2002; 2002US-00087192.
XX (SAGR-) SAGRES DISCOVERY.
XX Morris DW;
XX MPI; 2003-328604/31.
XX Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
XX comprises a nucleotide sequence.
XX Claim 1; SEQ ID NO 529; Opp; English.
XX The present invention relates to novel DNA and protein sequences which
XX are associated with carcinomas. The sequences are useful for: (i) for
XX screening drug candidates; (ii) for screening of bioactive agent capable
XX of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
XX a bioactive agent capable of modulating the activity of CAP; (iv) for
XX evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
XX carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
XX carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
XX (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
XX determining Carcinoma Associated (CA) gene copy number. In addition, the
XX CA genes are useful as DNA vaccines and the CAP are useful as markers of
XX carcinoma including lymphoma. The present sequence is one such CA coding
XX sequence. Note: This patent is an equivalent to basic patent
XX US2002182586A1, for which no sequence data was published
XX
SQ Sequence 211257 BP; 55148 A; 47354 C; 46616 G; 57484 T; 0 U; 4655 Other;
Query Match 14.4%; Score 72; DB 11; Length 211257;
Best Local Similarity 68.9%; Pred. No. 1.1e-09;
Matches 113; Conservative 0; Mismatches 50; Indels 1; Gaps 1;
QY 278 ATCTGCAAGTCCCTGAGTGGCAGAGGAGGACCCATGTGAGGCAAGAAATTCTGGC 337
DB 72685 ATCTGCGAGCCACCATGCGCATACAGATGAGATCTTGAGAGGCGCAAGATCTCTGGT 72626
QY 338 TCAACACAGCTTAGCTCCCTGCTGTGTTCAACTTTGAGAGTTTGACCAAGCACTT 397
DB 72625 TCAATTACAC-TAAATAGCTGAGAGTTGTTCAACATCGGAGAGTCTGATCTCAATAGTT 72567
QY 398 TATTTTGGACATATTTTAAACAGACACAACTTTGGAAAAAGTT 441
DB 72566 TATTTTGGACATATTTTAAAGCAGACATATTTGTTGCAATGTT 72523
RESULT 4
ACN43912
ID ACN43912 standard; DNA; 233060 BP.
XX
XX ACN43912;
XX 18-NOV-2004 (first entry)
XX Mouse genomic sequence mCG11034.
XX Cytostatic; carcinoma; lymphoma; cancer; murine; gene; ss.
XX Mus musculus.
XX OS

XX
PN WO2003073826-A2.
XX 12-SEP-2003.
XX 28-FEB-2003; 2003WO-US006235.
XX 01-MAR-2002; 2002US-00087192.
XX (SAGR-) SAGRES DISCOVERY.
XX Morris DW;
XX MPI; 2003-328604/31.
XX Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
XX comprises a nucleotide sequence.
XX Claim 1; SEQ ID NO 97; Opp; English.
XX The present invention relates to novel DNA and protein sequences which
XX are associated with carcinomas. The sequences are useful for: (i) for
XX screening drug candidates; (ii) for screening of bioactive agent capable
XX of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
XX a bioactive agent capable of modulating the activity of CAP; (iv) for
XX evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
XX carcinoma; (vi) for inhibiting the activity of CAP; (vii) as a biochip;
XX carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
XX (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
XX determining Carcinoma Associated (CA) gene copy number. In addition, the
XX CA genes are useful as DNA vaccines and the CAP are useful as markers of
XX carcinoma including lymphoma. The present sequence is one such CA coding
XX sequence. Note: This patent is an equivalent to basic patent
XX US2002182586A1, for which no sequence data was published
XX
SQ Sequence 233060 BP; 56260 A; 53970 C; 52927 G; 62345 T; 0 U; 7558 Other;
Query Match 13.8%; Score 69.2; DB 11; Length 233060;
Best Local Similarity 59.1%; Pred. No. 7.6e-09;
Matches 137; Conservative 0; Mismatches 93; Indels 2; Gaps 1;
QY 270 GTGGGATATCTGCAAGTGCATGATGCGAGGAGGACCCATGTGAGGCAAGAAAGA 329
DB 79572 GTAGGCTCCCTGCGAGCTCACCACAGCAAGGAAACCCACTTGAGGTGAAGATGA 79631
QY 330 ATTCTGCTCCACACAGCTTAGCTCCCTGCTGTGTTCAACTTTGAGAGTTGACAC 389
DB 79632 GCCCAGGTGACAGCACTTAGCAGGCCCTGG--GGCAGAAACCTCCAGAGTGTGACCTC 79689
QY 390 AAGCAGCTTATTTTGTGACATATTTAAACAGACACAACTTTGGGAAAAAGTTTCTATG 449
DB 79690 AGGGTTTACCTTTCTGCTCCTTAAACAGTAAACCTTTGAGGGAAGTCTCCACGTG 79749
QY 450 AAATATTCACAAATAAGCTTAAGGCGTACATTAATAATCCCTTGCA 501
DB 79750 GCATTTTACGAACAACAGCTTTCAAGGATAGTTTCATTTGAATCCCTTGCA 79801
RESULT 5
ABD32895
ID ABD32895 standard; DNA; 89824 BP.
XX
XX ABD32895;
XX 18-NOV-2004 (first entry)
XX Mouse cancer-associated genomic DNA MD18-037.
XX Mouse; da; cancer-associated protein; gene; cytostatic; cancer;
XX leukaemia; lymphoma; CAP.
XX Mus musculus.
XX OS

EN WO2004074320-A2.
 PD 02-SEP-2004.
 XX 17-FEB-2004; 2004WO-US004730.
 PF 14-FEB-2003; 2003US-00367094.
 XX 14-MAR-2003; 2003US-0038838.
 PR 15-APR-2003; 2003US-00417375.
 PR 13-JUN-2003; 2003US-00461862.
 PR 15-SEP-2003; 2003US-00663431.
 PR 15-DEC-2003; 2003US-00737318.
 XX (SAGR-) SAGRES DISCOVERY INC.
 PA Morris DW, Morrie DW, Malandro MS;
 PI WPI; 2004-652914/63.
 DR New isolated cancer-associated polynucleotides and polypeptides useful
 PT for diagnosing, preventing or treating cancers, especially lymphoma and
 PT leukemia, or in screening for agents that modulate cancer.
 XX disclosure; seqid 591; 310pp; English.
 PS The invention relates to an isolated nucleic acid comprising at least 10
 XX contiguous nucleotides of any of the 233 polynucleotide sequences given
 CC in the specification, or its complement. The nucleic acids encode cancer-
 CC associated proteins. Also included are an expression vector comprising
 CC the isolated nucleic acid cited above, a host cell comprising the above
 CC recombinant nucleic acid or expression vector, a microarray for detecting
 CC a cancer-associated (CA) nucleic acid comprising at least one probe
 CC comprising at least 10 contiguous nucleotides of any of the above-
 CC mentioned nucleotide sequences, an isolated polypeptide (encoded within
 CC an open reading frame of a CA sequence selected from any of the 95
 CC polynucleotide sequences as mentioned in the specification, or its
 CC complement), an isolated antibody, (or its antigen binding fragment) that
 CC binds to the above polypeptide, a hybridoma that produces the above
 CC monoclonal antibody, a pharmaceutical composition comprising the above
 CC antibody and a pharmaceutical excipient, a kit for detecting cancer
 CC cells (comprising the antibody cited above, methods for diagnosing cancer
 CC or for detecting the presence or absence of cancer cells in an
 CC individual, a method for inhibiting growth of cancer cells in an
 CC individual, a method for delivering a therapeutic agent to cancer cells
 CC in an individual, an electronic library comprising the above
 CC polynucleotide or polypeptide (or their fragments), methods of screening
 CC for anticancer activity or for a bioactive agent capable of modulating
 CC the activity of a CA protein (CAP), methods for detecting cancer
 CC associated with expression of a polypeptide in a test cell sample, a
 CC method for treating cancers and a method for inhibiting the expression of
 CC CA gene in a cell. The composition and methods are useful for detecting,
 CC diagnosing, preventing and treating cancers, especially lymphoma and
 CC leukemia. These may also be used in screening for agents that modulate
 CC cancer. The present sequence is a mouse CAP genomic sequence. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 89824 BP; 26577 A; 18285 C; 18254 G; 26575 T; 0 U; 133 Other;
 SQ
 Query Match 13.1%; Score 65.6; DB 13; Length 89824;
 Best Local Similarity 68.5%; Pred. No. 5.9e-08;
 Matches 137; Conservative 0; Mismatches 54; Indels 9; Gaps 3;
 QY 272 GGGGATATCTGCAAGTCCATGAGTGGCAGAGGAGCAAGCCATGTGAGGCAAGAGAAAT 331
 DB 3139 GGGGATCTCTGCAAAATTCATGTGTCAGAGAGTTTGAAGC-TTGAAGGCTATTAAGACTA 3197
 QY 332 TCTGGCTCAACAGACT-----TAGCTCCCTGGTGTGTGTTCAAACTTTGAGAGTTTG 384
 DB 3198 TCCAGTTCAACAGAGACTCAGTAAGTAATATCAAGTGTGTGTTCAAACTTTCTGAGAGCTTG 3257
 QY 385 ACAGCAAGCACTTTATTTTGAAGTATTTAAAGAGCAAGCAACTTTGG-GAAAAAGTTT 443

Db 3258 ACCTCAGGTAGTTATGTGAGCAGCATTTTAAGCAGCAGCAATTTTGGAAAAAGTTTC 3317
 QY 444 CTTATGAAAATTAATCAAT 463
 Db 3318 CTGATGACAAATTCATCAAT 3337
 RESULT 6
 WP Sequence split into 7 fragments LOCUS ABX16390 Accession Abx16390
 WP ABX16390_0 Begin End
 WP ABX16390_1 1 110000
 WP ABX16390_2 100001 210000
 WP ABX16390_3 200001 310000
 WP ABX16390_4 300001 410000
 WP ABX16390_5 400001 510000
 WP ABX16390_6 500001 610000
 WP ABX16390_7 600001 659158
 ID ABX16390 standard; DNA; 659158 BP.
 AC ABX16390;
 XX 09-APR-2003 (first entry)
 DT Mouse high growth region.
 DE High growth region; high growth phenotype; Soc2; body size;
 KW suppressor of cytokine signaling 2; ds; mouse.
 OS Mus sp.
 XX US2002155564-A1.
 PN 24-OCT-2002.
 PD 26-JAN-2001; 2001US-00771208.
 PF 29-DEC-1997; 97US-00999477.
 PR (REGC) UNIV CALIFORNIA.
 PA Medrano JF, Bradford E, Horvat S;
 PI WPI; 2003-182637/18.
 DR Novel gene that when downregulated or knocked-out, results in high growth
 PT phenotype, useful for regulating body size in mammals e.g. rodent, bovine
 PT and canine.
 XX
 XX Disclosure; SEQ ID NO 18; 49pp; English.
 PS The invention describes an isolated nucleic acid molecule encoding a gene
 CC product that, when knocked out, results in a high growth (hg) phenotype.
 CC For example a nucleic acid disrupting the Soc2 gene is useful for
 CC producing an animal characterised by a hg phenotype, by inhibiting
 CC expression of Soc2 (suppressor of cytokine signaling 2) gene. The nucleic
 CC acids of the invention are useful for regulating body size in mammals.
 CC gene. The nucleic acids of the invention are useful for regulating body
 CC size in mammals. This sequence represents the mouse high growth region.
 CC Note: This sequence did not form part of the printed specification but
 CC was obtained in electronic format directly from the US patent office at
 CC seqdata.uspto.gov/sequence.html?docid=20020155564
 XX
 XX Sequence 659158 BP; 177521A; 147222C; 149416G; 184754T; 0U; 2470Other;
 SQ
 Query Match 13.0%; Score 65.2; DB 8; Length 110000;
 Best Local Similarity 73.8%; Pred. No. 8.5e-08;
 Matches 96; Conservative 0; Mismatches 33; Indels 1; Gaps 1;
 QY 317 GAGGCAAGAGAAATTTGCTCAACAGACTTAGCTCCCTGGTGTGTTCAAACTTTG 376
 Db 14497 GAGGCAAGAGAAAGTCTAGTTCAACAGACTCAG-TAGCTCATGTGTTCAAACTTTG 14439

QY 467 GCTTAAGCATGACTACATTAAATGCTTTGCAA 501
 DB 62016 GCTTTAGCATGGCTACATCAAACTCTTTGTAA 61982

RESULT 11
 ADB72539/c
 ID ADB72539 standard; DNA; 67832 BP.

AC ADB72539;

DT 04-DEC-2003 (first entry)

DE Mouse Itk gene.

KM mouse; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;
 KM cancer; neoplasm; adenocarcinoma; sarcoma; gene.

OS Mus sp.

PN WO2003008583-A2.

XX 30-JAN-2003.

PF 26-DEC-2001; 2001WO-US051291.

XX 02-MAR-2001; 2001US-00798586.

PR 23-OCT-2001; 2001US-00004113.

PR 08-NOV-2001; 2001US-00052482.

PR 30-NOV-2001; 2001US-00997722.

PR 20-DEC-2001; 2001US-00034650.

XX (SAGR-) SAGRES DISCOVERY.

XX PI Morris DW, Engelhard EK;

XX WPI; 2003-239337/23.

XX New recombinant nucleic acid, useful for treating carcinomas, lymphomas,
 PT cancers, neoplasm, adenocarcinoma, or sarcomas.

PS Claim 1; SEQ ID NO 367; 2304pp; English.

XX The invention relates to a novel recombinant nucleic acid comprising a
 CC nucleotide sequence selected from any of the 660 sequences fully defined
 CC in the specification. A polynucleotide of the invention has cytostatic
 CC activity, and may have a use in gene therapy, or in a vaccine. The
 CC recombinant nucleic acids and polypeptides are useful for treating
 CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
 CC sarcomas. The present sequence represents a mouse gene of the invention.

XX Sequence 67832 BP; 17054 A; 15504 C; 15824 G; 18278 T; 0 U; 1172 Other;

SO Query Match 12.3%; Score 61.4; DB 10; Length 67832;

Best Local Similarity 77.9%; Pred. No. 9.2e-07;

Matches 74; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 407 CATATTTAAACAGACAACTTTGGGAAAAGTTTCTTATGAAATTAATCAATTA 466

DB 62076 CACATTTAAACAGACAACTTTGGGAAAAGTTTCTTATGAAATTAATCAATTA 62017

QY 467 GCTTAAGCATGACTACATTAAATGCTTTGCAA 501

DB 62016 GCTTTAGCATGGCTACATCAAACTCTTTGTAA 61982

RESULT 12

ADCB5281/c

XX ID ADCB5281 standard; DNA; 67832 BP.

XX AC ADCB5281;

DT 01-JAN-2004 (first entry)

XX Mouse Itk genomic sequence.

XX Cytostatic; gene therapy; vaccine; cancer; carcinoma-associated gene; CA;
 KM secreted; transmembrane; intracellular; ds.

XX Mus sp.

XX WO2003045230-A2.

XX 05-JUN-2003.

PF 02-DEC-2002; 2002WO-US038582.

XX 30-NOV-2001; 2001US-00997722.

XX (SAGR-) SAGRES DISCOVERY.

XX PI Morris DW, Engelhard EK;

XX WPI; 2003-513603/48.

XX New recombinant nucleic acid comprising a nucleotide sequence of any of
 PT the carcinoma-associated (CA) genes, useful for screening for drug
 PT candidates for diagnosing or treating carcinomas.

XX Claim 1; SEQ ID NO 67; 983pp; English.

XX The invention relates to a recombinant nucleic acid comprising a
 CC nucleotide sequence selected from any of the fully defined carcinoma-
 CC associated (CA) genes from the 50 tables given in the specification. The
 CC CA proteins are secreted, transmembrane or intracellular proteins. The
 CC recombinant nucleic acids are useful for screening for drug candidates
 CC for diagnosing or treating carcinomas. Sequences given in ADCB5215-
 CC ADCB5514 represent CA genes of the invention.

XX Sequence 67832 BP; 17054 A; 15504 C; 15823 G; 18278 T; 0 U; 1173 Other;

SO Query Match 12.3%; Score 61.4; DB 10; Length 67832;

Best Local Similarity 77.9%; Pred. No. 9.2e-07;

Matches 74; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 407 CATATTTAAACAGACAACTTTGGGAAAAGTTTCTTATGAAATTAATCAATTA 466

DB 62076 CACATTTAAACAGACAACTTTGGGAAAAGTTTCTTATGAAATTAATCAATTA 62017

QY 467 GCTTAAGCATGACTACATTAAATGCTTTGCAA 501

DB 62016 GCTTTAGCATGGCTACATCAAACTCTTTGTAA 61982

RESULT 13

ADM74396/c

XX ID ADM74396 standard; DNA; 67832 BP.

XX AC ADM74396;

XX 01-JUL-2004 (first entry)

XX Murine carcinoma associated (CA) nucleic acid #34.

XX Mouse; carcinoma associated nucleic acid; CA nucleic acid; gene; ds;
 KM carcinoma associated protein; CAP; carcinoma; leukemia; lymphoma;
 KM cytostatic.

XX Mus musculus.

XX US2004072154-A1.

XX 15-APR-2004.

XX 30-NOV-2001; 2001US-00997722.

XX	22-DEC-2000;	2000US-00747377.	
PR	02-MAR-2001;	2001US-00796586.	
PA	(MORR/) MORRIS D W.		
PA	(ENGE/) ENGELHARD E K.		
PI	Morris DW, Engelhard EK;		
DR	WPI; 2004-328562/30.		
XX			
XX	New carcinoma associated gene or protein, useful for preparing a		
PT	composition for diagnosing or treating carcinoma e.9., leukemia or		
PT	lymphoma.		
PS	Claim 1; SEQ ID NO 67; 29pp; English.		
XX			
XX	The invention relates to new recombinant nucleic acids. The invention		
CC	also relates to a host cell comprising a recombinant nucleic acid or		
CC	expression vector, an expression vector comprising a recombinant nucleic		
CC	acid, a recombinant protein, a method of screening for drug candidates, a		
CC	method of screening for a bioactive agent capable of binding to a		
CC	carcinoma associated protein (CAP) encoded by a nucleotide sequence, a		
CC	method of screening for a bioactive agent capable of modulating the		
CC	activity of a CAP, a method of evaluating the effect of a candidate		
CC	carcinoma drug, a method of diagnosing carcinoma, a method for inhibiting		
CC	the activity of a CAP, a method of treating carcinomas, a method of		
CC	neutralising the effect of a CAP and a method of diagnosing carcinoma or		
CC	propensity to carcinomas. A method of evaluating the effect of a candidate		
CC	carcinoma drug comprises administering the drug to a patient, removing a		
CC	cell sample from the patient and determining alterations in the		
CC	expression or activation of a gene comprising the nucleotide sequence. A		
CC	method of diagnosing carcinoma comprises determining the expression of		
CC	one or more genes comprising the nucleic acid sequence in a first tissue		
CC	type of a first individual and comparing the expression of the gene from		
CC	a second normal tissue type from the first individual or a second		
CC	affected individual, where a difference in the expression indicates		
CC	that the first individual has carcinoma. A method of inhibiting the		
CC	activity of a CAP comprises binding an inhibitor to the CAP. Treating		
CC	carcinomas comprises administering to a patient an inhibitor of CAP.		
CC	Neutralising the effect of a CAP comprises contacting an agent specific		
CC	for the CAP. The polypeptide specifically binds to the protein encoded by		
CC	the nucleic acid. It comprises an antibody that specifically binds to the		
CC	protein encoded by the nucleic acid. The nucleic acids are useful for		
CC	preparing a composition for diagnosing or treating carcinoma e.9.,		
CC	leukaemia or lymphoma. This sequence represents a murine carcinoma		
CC	associated (CA) nucleic acid of the invention. Note: the sequence data		
CC	for this patent did not form part of the printed specification but was		
CC	obtained in electronic format directly from USPTO at		
CC	seqdata.uspto.gov/sequence.html.		
XX			
SQ	Sequence 67832 BP; 17054 A; 15504 C; 15824 G; 18278 T; 0 U; 1172 Other;		
XX			
XX	Query Match	12.3%; Score 61.4; DB 12; Length 67832;	
XX	Best Local Similarity	77.9%; Pred. No. 9.2e-07;	
XX	Matches	74; Conservative 0; Mismatches 21; Indels 0; Gaps 0;	
OY	407 CATATTTAACAGGACCAACAACCTTTGGGAAAAAGTTTCTATGAAAAATTATGACAAATAA	466	
DB	62076 CACATTTTAAACACAGTAAGACCTTTGGGAAAAAGTTTCTTGTGACGTGATCAGGCAAA	62017	
OY	467 GCTTAAGCATGACTACATTAATAAATGCTTTGCAA	501	
DB	62016 GCTTTAGGCGATGCTACATCAAAACCTCTTTTGTAA	61982	
XX			
XX	RESULT 14		
AC	ACN44040		
XX	ACN44040 standard; DNA; 210528 BP.		
XX	ACN44040;		
DT	18-NOV-2004 (first entry)		

[illegible]

OS Mus musculus.
 XX WO2004058146-A2.
 XX 15-JUL-2004.
 PD 15-DEC-2003; 2003WO-US040081.
 XX 17-DEC-2002; 2002US-00322281.
 PR (SAGR-) SAGRES DISCOVERY INC.
 XX (SAGR-) SAGRES DISCOVERY INC.
 PA Morris DW, Malandro MS;
 XX WPI; 2004-499109/47.
 DR Novel human cancer associated protein encoded within open reading frame
 XX of cancer associated gene, useful as targets for diagnosing cancer.
 PT
 XX
 XX Disclosure; SEQ ID NO 587; 182bp; English.
 PS
 XX
 CC The invention relates to cancer-associated proteins (CAP) and the cancer-
 CC associated (CA) nucleic acids encoding them. The invention also relates
 CC to a method for treating cancers involving administering to a patient an
 CC inhibitor of CAP, and a method of screening for anticancer activity in a
 CC potential drug involving providing a cell that expresses a CA gene,
 CC contacting a tissue sample derived from a cancer cell with an anticancer
 CC drug candidate and monitoring the effect of the anticancer drug candidate
 CC on expression of the CA gene. The CAP proteins are useful for detecting
 CC cancer associated with expression of a CAP protein in a test cell sample
 CC and for screening for a bioactive agent capable of modulating the
 CC activity of a CAP protein. The CA nucleic acids are useful for diagnosing
 CC cancer, involving determining the expression of a CA nucleic acid in a
 CC tissue. This sequence represents a murine CA gene of the invention. Note:
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 121129 BP; 34090 A; 24248 C; 25630 G; 36418 T; 0 U; 743 Other;
 Query Match 12.1%; Score 60.4; DB 13; Length 121129;
 Best Local Similarity 71.3%; Pred. No. 2.4e-06;
 Matches 107; Conservative 0; Mismatches 41; Indels 2; Gaps 2;
 QY 314 TGTAGGCAAGAGGATTCGCTCAACAGCTTAGCTCCCTGTGTGTGTTCAACT 373
 Db 119057 TGAGAGCAAAAGGGCGATTGTTCAACATGACTCA-ATAGCTGGTGTAGTTCAATT 118999
 QY 374 TTGAGAGTTGACCAACAGCACTTTATTTTGTGACATATTTAAACAGACCAACTTTGGG 433
 Db 118998 TTGAAAGCTGACCTCAAGTATTTTATTTACGGCATTTTAAAGCAGCAGAA-ATTGAA 118940
 QY 434 AAAAAGTTTCTATGAAATTAATCAAT 463
 Db 118939 AAGAAAGTTGTAAATGATTAATAGCTTAAT 118910

Search completed: February 10, 2005, 09:41:03
 Job time : 345.632 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 22:08:58, Search time 103.92 Seconds
(without alignments)
7888.487 Million cell updates/sec

Title: US-09-926-375b-7_COPY_5000_5500

Perfect score: 501

Sequence: 1 ggtcagccatgtcttatt.....acattcaaatgcttgcga 501

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Issued Patents NA:*

- 1: /cgn2_6/prodata/1/ina/5A_COMB.seq:*
- 2: /cgn2_6/prodata/1/ina/5B_COMB.seq:*
- 3: /cgn2_6/prodata/1/ina/6A_COMB.seq:*
- 4: /cgn2_6/prodata/1/ina/6B_COMB.seq:*
- 5: /cgn2_6/prodata/1/ina/PCTUS_COMB.seq:*
- 6: /cgn2_6/prodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	37.6	7.5	144362	US-09-949-016-16066	Sequence 16066, A
2	35.8	7.1	15027	US-09-949-016-12660	Sequence 12660, A
3	35.8	7.1	15036	US-09-949-016-13351	Sequence 13351, A
4	35.6	7.1	1053	US-09-710-279-3127	Sequence 3127, Ap
5	35.6	7.1	1149	US-09-134-001C-2612	Sequence 2612, Ap
6	35.6	7.1	1492	US-09-710-279-3677	Sequence 3677, Ap
7	35.6	7.1	3825	US-09-710-279-3794	Sequence 3794, Ap
8	34.8	6.9	7218	US-08-232-463-14	Sequence 14, Appl
9	34.8	6.9	10543	US-09-949-016-13138	Sequence 13138, A
10	34.4	6.9	101472	US-09-949-016-15861	Sequence 15861, A
11	34.4	6.9	102738	US-09-949-016-12447	Sequence 12447, A
12	34.4	6.9	192700	US-09-949-016-11820	Sequence 11820, A
13	34.4	6.9	192704	US-09-949-016-17182	Sequence 17182, A
14	34.2	6.8	600	US-09-621-976-14175	Sequence 14175, A
15	34	6.8	118382	US-09-949-016-15996	Sequence 15996, A
16	34	6.8	118382	US-09-949-016-15997	Sequence 15997, A
17	33.8	6.7	1141	US-09-806-708B-22	Sequence 22, Appl
18	33.8	6.7	86439	US-09-949-016-11945	Sequence 11945, A
19	33.8	6.7	86440	US-09-949-016-16990	Sequence 16990, A
20	33.6	6.7	7005	US-09-949-016-2891	Sequence 2891, Ap
21	33.6	6.7	7005	US-09-949-016-2892	Sequence 2892, Ap
22	33.6	6.7	7005	US-09-949-016-2893	Sequence 2893, Ap
23	33.6	6.7	7005	US-09-949-016-2894	Sequence 2894, Ap
24	33.6	6.7	57280	US-09-949-016-11796	Sequence 11796, A
25	33.6	6.7	57280	US-09-949-016-12843	Sequence 12843, A
26	33.6	6.7	57280	US-09-949-016-12844	Sequence 12844, A
27	33.6	6.7	57280	US-09-949-016-12846	Sequence 12846, A

28	33.6	6.7	57280	US-09-949-016-13542	Sequence 13542, A
29	33.6	6.7	57280	US-09-949-016-13543	Sequence 13543, A
30	33.6	6.7	57280	US-09-949-016-13544	Sequence 13544, A
31	33.6	6.7	57280	US-09-949-016-13545	Sequence 13545, A
32	33.6	6.7	57280	US-09-949-016-14633	Sequence 14633, A
33	33.6	6.7	57280	US-09-949-016-14634	Sequence 14634, A
34	33.6	6.7	57280	US-09-949-016-14635	Sequence 14635, A
35	33.6	6.7	57280	US-09-949-016-14636	Sequence 14636, A
36	33.6	6.7	57280	US-09-949-016-14637	Sequence 14637, A
37	33.6	6.7	57280	US-09-949-016-14638	Sequence 14638, A
38	33.6	6.7	57280	US-09-949-016-14639	Sequence 14639, A
39	33.6	6.7	57280	US-09-949-016-14640	Sequence 14640, A
40	33.6	6.7	57280	US-09-949-016-13768	Sequence 13768, A
41	33.6	6.7	116592	US-09-818-512-3	Sequence 3, Appl1
42	33.6	6.7	152132	US-09-949-016-13845	Sequence 13845, A
43	33.6	6.7	152145	US-09-949-016-12371	Sequence 12371, A
44	33.4	6.7	1664976	US-08-916-421B-1	Sequence 1, Appl1
45	33.4	6.7	1664976	US-09-692-570-1	Sequence 1, Appl1

ALIGNMENTS

```

RESULT 1
; Sequence 16066, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0.
; SEQ ID NO 16066
; LENGTH: 144362
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(144362)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16066

Query Match 7.5%; Score 37.6; DB 4; Length 144362;
Best Local Similarity 54.5%; Pred. No. 2.1;
Matches 96; Conservative 0; Mismatches 79; Indels 1; Gaps 1;

QY 81 AGTCCAGGCTGACAGTGTGATACACAGGACGATGAGGCTCCACGCTGAACAG 140
DB 12004 AGCTTATCTCTGAGATGTGATGACAGAAAGAGGAGGACAGTCTCCAGAGGAA 11945
QY 141 TCAAGCTGGAGAGAGAAAGACGACACATCTCTTAACCAACATATGCTTGAATA 200
DB 11944 ACGAGCTCAGAAAG-GAATAAGGGGCAAAAAGTAGATCATGTTTATTTCCCAAT 11886
QY 201 CAACATATATATATATATATATATATATATATATATATATATATATATATATATAT 256
DB 11885 AACCTACATATATATATATATATATATATATATATATATATATATATATATATAT 11830

RESULT 2
; Sequence 12660, Application US/09949016
; Patent No. 6812339

```

```

RESULT 3
US-09-949-016-13351
; Sequence 13351, Application US//09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US//09/949, 016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241, 755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237, 768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231, 498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for windows Version 4.0
; SEQ ID NO 13351
; LENGTH: 15036
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13351

```

Query Match	7.1%	Score 35.6	DB 3	Length 1149
Best Local Similarity	55.7%	Pred No. 0.55		
Matches 68; Conservative	0	Mismatches 54	Indels 0	Gaps 0

```

OY 369 AAACCTTGAGAGTTTGACCAACAAGCACTTATTATTGACATATTATAACAGAGCAACT 428
Db 497 AATCTTTCACAGTTGATCTGAAGAATATGATTTAAGTCATTTGAAACCGAGTCATATCA 556
OY 429 TTGGGAAAAGTTTCTTATGAAAATATATACATATAAGTTAAGGCATGACTATTA 488
Db 557 TTACCGTAGATGTTTCTAGTGAAGAAAGTTACCATATAGAGTAAACGTATTAATCA 616
OY 489 AA 490
Db 617 TA 618

RESULT 6
US-09-710-279-3677/c
; Sequence 3677, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3677
; LENGTH: 1492
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-09-710-279-3677

Query Match 7.1%; Score 35.6; DB 4; Length 1492;
Beet Local Similarly 55.7%; Pred. No. 0.64;
Matches 68; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

OY 369 AAACCTTGAGAGTTTGACCAACAAGCACTTATTATTGACATATTATAACAGAGCAACT 428
Db 1411 AATCTTTCACAGTTGATCTGAAGAATATGATTTAAGTCATTTGAAACCGAGTCATATCA 1352
OY 429 TTGGGAAAAGTTTCTTATGAAAATATATACATATAAGTTAAGGCATGACTATTA 488
Db 1351 TTACCGTAGATGTTTCTAGTGAAGAAAGTTACCATATAGAGTAAACGTATTAATCA 1292
OY 489 AA 490
Db 1291 TA 1290

RESULT 7
US-09-710-279-3794
; Sequence 3794, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3794
; LENGTH: 3825
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic

```

```

; OTHER INFORMATION: nucleic acid sequence
; US-09-710-279-3794

Query Match          7.1%; Score 35.6; DB 4; Length 3825;
Best Local Similarity 5.7%; Pred. No. 1.1;
Matches 66; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

Oy      369  AAACCTTGAGAGTTTGACCAACAGCACTTATTTTGGACATATTAAACGAGCACT 428
Db      1922  AATCTTCACAGTTGAATCTGAAGAATATGATTTAAGTCATTCGAAACCGGTCATACA 1981

Oy      429  TTGGGAAAAGCTTCTTATGAAAATTATGACATAAAGCTTAAGCATGACTATTAA 488
Db      1982  TTACCGTAGATGTTCTTAGTGAAAAAGTTACCATATAGCTAAACGTCATTATTCATCA 2041

Oy      489  AA 490
Db      2042  TA 2043

RESULT 8
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: PTZgpc-Fls
; US-08-232-463-14

Query Match          6.9%; Score 34.8; DB 1; Length 7218;
Best Local Similarity 5.7%; Pred. No. 2.9;
Matches 18; Conservative 164; Mismatches 136; Indels 0; Gaps 0;

```


Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11820
LENGTH: 192700
TYPE: DNA
ORGANISM: Human
US-09-949-016-11820

Query Match 6.9%; Score 34.4; DB 4; Length 192700;
Best Local Similarity 63.1%; Pred. No. 27;
Matches 53; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
Qy 163 CCAGCAGACATTCCTTCAACCACTATGCTTGAACCAACATATTATATACATATA 222
Db 167541 CCACATCAGACCTCCCTAAGCGCCATGCTTGAACCATGATATTATATACCTATA 167482
Qy 223 TTGCATTATGAGACAGCTAAAT 246
Db 167481 TTTAATTATCATTTGGTTAAAT 167458

RESULT 13
US-09-949-016-17182/c
Sequence 17182, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 17182
LENGTH: 192704
TYPE: DNA
ORGANISM: Human
US-09-949-016-17182

Query Match 6.9%; Score 34.4; DB 4; Length 192704;
Best Local Similarity 63.1%; Pred. No. 27;
Matches 53; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
Qy 163 CCAGCAGACATTCCTTCAACCACTATGCTTGAACCAACATATTATATACATATA 222
Db 167541 CCACATCAGACCTCCCTAAGCGCCATGCTTGAACCATGATATTATATACCTATA 167482
Qy 223 TTGCATTATGAGACAGCTAAAT 246
Db 167481 TTTAATTATCATTTGGTTAAAT 167458

RESULT 14
US-09-621-976-14175/c
Sequence 14175, Application US/09621976
Patent No. 6639063
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Jobert, S.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: ESTs and Encoded Human Proteins.
FILE REFERENCE: GENSET.054PR2
CURRENT APPLICATION NUMBER: US/09/621,976
CURRENT FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 19335
SOFTWARE: Patent.pm
SEQ ID NO 14175
LENGTH: 600
TYPE: DNA
ORGANISM: Homo sapiens
US-09-621-976-14175

Query Match 6.8%; Score 34.2; DB 4; Length 600;
Best Local Similarity 49.7%; Pred. No. 1.1;
Matches 87; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
Qy 280 CTGCAGATGCATGAGTGGCAGAGGACAGCCATGTGAGGCAAGAGATTTGAGCTC 339
Db 494 CTGCCAGGATTTGACAGGAGAGGAGATGATAGTGAGGCAAGGATTTTAGGCGC 435
Qy 340 AACACAGCTTACGCTCCGTGTGTTGTTCAACTTTGAGAGTTGACACACACTTTA 359
Db 434 AGTGAAGTGTGAGTGTACAGTGTGGTATCATCTATACATTTGCAATGTATCTGA 375
Qy 400 TTTTATCATATTATAACAGACACCACTTTGGAAAAAGTTTCTATGAAAT 454
Db 374 ATATACACAAAGACAAACCTAACAACTCCGACTTGGCTGATATATATAT 320

RESULT 15
US-09-949-016-15996
Sequence 15996, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 15996
LENGTH: 118382
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(118382)
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15996

Query Match 6.8%; Score 34; DB 4; Length 118382;
Best Local Similarity 51.3%; Pred. No. 27;
Matches 79; Conservative 0; Mismatches 75; Indels 0; Gaps 0;
Qy 112 GCAGCATGAGGTCTCAGCTGAGGAGTCAAGCTGGCAGAGAGAAAGACACGACAC 171
Db 79717 GCACTTTGAGAGGCGAAGGAGAGATGCTTGAGACCAAGAGTTGAGACCGAGCTGG 79776

OY 172 ATTCTTCAACCACTAGTCTTGAAACAACATATTATATGACATATATGCAATTTA 231
 Db 79777 GTAACATAGACCTTGTCTTTATTAATAAATAAGTAATAATAGTACATGTATATAATTAAGA 79836
 OY 232 TGAGACAGCTAAATGTACTCGGGTAGCATGACT 265
 Db 79837 TTAAGGAAGTATATATTAAGTGTAGCAGTAAT 79870

Search completed: February 10, 2005, 17:42:49
 Job time : 107.92 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 22:13:58 ; Search time 365.063 Seconds
(without alignments)
8097.594 Million cell updates/sec

Title: US-09-926-375b-7_COPY_5000_5500

Perfect score: 501 1 ggcagcccatgctctattc.....acatcaaatgccttgcaa 501

Scoring table: IDENTITY WUC
Gapop 10.0, Gapext 1.0

Searched: 5378673 seqs, 2950229984 residues

Total number of hits satisfying chosen parameters: 10757346

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

```
1: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/prodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/prodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/prodata/2/pubpna/US06_NEW_PUB.seq:*
5: /cgn2_6/prodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/prodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/prodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/prodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/prodata/2/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/prodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/prodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/prodata/2/pubpna/US09C_PUBCOMB.seq:*
13: /cgn2_6/prodata/2/pubpna/US10_PUBCOMB.seq:*
14: /cgn2_6/prodata/2/pubpna/US10C_PUBCOMB.seq:*
15: /cgn2_6/prodata/2/pubpna/US10_PUBCOMB.seq:*
16: /cgn2_6/prodata/2/pubpna/US10_PUBCOMB.seq:*
17: /cgn2_6/prodata/2/pubpna/US10_PUBCOMB.seq:*
18: /cgn2_6/prodata/2/pubpna/US10_PUBCOMB.seq:*
19: /cgn2_6/prodata/2/pubpna/US10_NEW_PUB.seq:*
20: /cgn2_6/prodata/2/pubpna/US11_NEW_PUB.seq:*
21: /cgn2_6/prodata/2/pubpna/US60_NEW_PUB.seq:*
22: /cgn2_6/prodata/2/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	72	14.4	211257	13	US-10-087-192-529
C 2	69.2	13.8	233060	13	US-10-087-192-97
C 3	65.2	13.0	659158	9	US-09-771-108-20
C 4	64.2	12.8	225883	14	US-10-175-523-57
C 5	63.6	12.7	48133	13	US-10-087-192-1519
C 6	62.4	12.5	47884	17	US-10-085-117-121
C 7	61.8	12.3	105077	18	US-10-417-715-15
C 8	61.6	12.3	145068	18	US-10-322-281-33
C 9	61.4	12.3	67832	11	US-09-997-722-67
C 10	60.8	12.1	210528	13	US-10-087-192-289
C 11	60.4	12.1	121129	18	US-10-322-281-587

12	60	12.0	421609	18	US-10-367-094-122	Sequence 122, App
13	58.6	11.7	96599	11	US-09-997-722-199	Sequence 199, App
14	58	11.6	33488	17	US-10-085-117-235	Sequence 235, App
15	58	11.6	75101	13	US-10-087-192-1337	Sequence 1237, App
16	57.2	11.4	11141	14	US-10-252-340-16	Sequence 16, App1
17	54.4	10.9	52256	13	US-10-087-192-1399	Sequence 1399, App
18	53.2	10.6	55996	11	US-09-997-722-7	Sequence 7, App11
19	50	10.0	247461	18	US-10-322-281-103	Sequence 103, App
C 20	49.6	9.9	23889	18	US-10-322-281-103	Sequence 9, App1
C 21	49.6	9.9	23889	18	US-10-417-375-9	Sequence 211, App1
C 22	49.2	9.8	96289	18	US-10-322-281-211	Sequence 186, App
C 23	47.8	9.5	52637	18	US-10-367-094-186	Sequence 1303, App
C 24	47.4	9.5	54552	13	US-10-087-192-1103	Sequence 10, App1
C 25	46.6	9.3	173808	13	US-10-003-806-10	Sequence 54369, A
C 26	38.6	7.7	4428	18	US-10-367-094-23	Sequence 23, App1
C 27	38.6	7.7	4428	18	US-10-367-094-23	Sequence 1393, App
C 28	37.8	7.5	27920	13	US-10-087-192-193	Sequence 769, App
C 29	37.6	7.5	50781	13	US-10-087-192-769	Sequence 210982,
C 30	37.2	7.4	571	13	US-10-027-632-210982	Sequence 93, App1
C 31	37.2	7.4	571	17	US-10-027-632-210982	Sequence 5729, App
C 32	36.4	7.3	100685	18	US-10-388-838-93	Sequence 1, App11
C 33	36.4	7.3	403035	18	US-10-741-601-5729	Sequence 33395, A
C 34	36.2	7.2	1691139	14	US-10-067-514-1	Sequence 173, App
C 35	36.2	7.2	1691139	17	US-10-425-115-33395	Sequence 37, App1
C 36	36	7.2	281	18	US-10-191-803-173	Sequence 4253, App
C 37	36	7.2	5042	17	US-10-034-650-37	Sequence 21385, A
C 38	36	7.2	78028	17	US-10-425-114-4252	Sequence 161951,
C 39	35.8	7.1	2171	17	US-10-425-114-21385	Sequence 3733, App
C 40	35.8	7.1	2400	17	US-10-425-114-21385	Sequence 2605, App
C 41	35.8	7.1	3579	18	US-10-425-115-161951	Sequence 28, App1
C 42	35.6	7.1	411	9	US-09-983-965-3733	Sequence 22, App1
C 43	35.4	7.1	929	10	US-09-887-527-28	
C 44	35.4	7.1	929	10	US-09-887-527-28	
C 45	35.4	7.1	2460	14	US-10-050-704-22	

ALIGNMENTS

```
RESULT 1
US-10-087-192-529/C
; Sequence 529, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 529
; LENGTH: 211257
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(211257)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-529

Query Match 14.4%; Score 72; DB 13; Length 211257;
Best Local Similarity 68.9%; Pred. No. 2,7e-09;
Matches 113; Conservative 0; Mismatches 50; Indels 1; Gaps 1;

CY 278 ATCTGCAAGTGCATGATGTCGACAGAGGACGACCATGTGAGCAAGAAATTCTGCG 337
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```


Db 72685 ATTCGACCCACCATGCGCATCAGTAAGATCTTGAAGAGGACAGAGATCTCTGT 72626
QY 338 TCACACAGCTTAGCTCCCTGGTGTGTTGTTCAAACTTTGAGAGTTGACCAAGCACTT 397
Db 72625 TCACACAGCTTAGCTCCCTGGTGTGTTGTTCAAACTTTGAGAGTTGACCAAGT 72567
QY 398 TATTTTGAATATTTTAAACAGACAGCAACTTTGGGAAAAAGTT 441
Db 72566 TATTTTGAATATTTTAAACAGACAGCAACTTTGGGAAAAAGTT 72523

RESULT 2
US-10-087-192-97
; Sequence 97, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE OF INVENTION: CANCER
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; PRIOR FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 97
; LENGTH: 233060
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(233060)
; OTHER INFORMATION: n = A,T,C or G
; US-10-087-192-97

Query Match 13.8%; Score 69.2; DB 13; Length 233060;
Best Local Similarity 59.1%; Pred. No. 1.9e-08;
Matches 137; Conservative 0; Mismatches 93; Indels 2; Gaps 1;

QY 270 GTGGGATATCTGCAAGTGCCTGATGTCAGAGGAGGACCCAAATGTGAGGCAAGAGA 329
Db 79572 GTGAGGGTCCCTGCGCCTCACCACCAAGGAGGAGCCCACTGAGGTGAAGATGA 79631
QY 330 ATTCTGCTCAACAGAGCTTAGCTCCCTGGTGTGTTCAAACTTTGAGAGTTGACCAAC 389
Db 79632 GCCCAGGTCAAGAGCACTTAGAGAGGCCCTGG--GGCAGAAACCTCCAGATGTGACCTC 79689
QY 390 AAGCATTATTTTATTTGACATATTTTAAACAGAGCAACTTTGGGAAAAAGTTTCTTATG 449
Db 79690 AAGGTTTACCTTTCTGCTCCTCAATAAAAGTAAACCTTTGAGGAAAGTCTCCACGIG 79749
QY 450 AAAATATATCAATAAAGCTTAGAGCATGACTCATTTAAATCCCTTTGAA 501
Db 79750 GCAATTATCAGAACAGCTTTCAGGAGTACTTTCATTGAACCTCCCTTGCA 79801

RESULT 3
US-09-771-208-20/c
; Sequence 20, Application US/09771208
; Patent No. US2002015564A1
; GENERAL INFORMATION:
; APPLICANT: MEDRANO, JUAN
; APPLICANT: BRADFORD, ERIC
; APPLICANT: HORVAT, SIMON
; TITLE OF INVENTION: CLONING OF A HIGH-GROWTH GENE
; FILE REFERENCE: 407T-923710US
; CURRENT APPLICATION NUMBER: US/09/771,208
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: US 08/999,477

PRIOR FILING DATE: 1997-12-29
NUMBER OF SEQ ID NOS: 20
SOFTWARE: PatentIn version 3.0
SEQ ID NO 20
LENGTH: 659158
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: misc feature
LOCATION: (123459)..(123478)
OTHER INFORMATION: n is unidentified a, c, g, or t
NAME/KEY: misc feature
LOCATION: (602466)..(602485)
OTHER INFORMATION: n is unidentified a, c, g, or t
NAME/KEY: misc feature
LOCATION: (546998)..(547017)
OTHER INFORMATION: n is unidentified a, c, g, or t
NAME/KEY: misc feature
LOCATION: (494715)..(494814)
OTHER INFORMATION: n is unidentified a, c, g, or t
NAME/KEY: misc feature
LOCATION: (390986)..(391005)
OTHER INFORMATION: n is unidentified a, c, g, or t
NAME/KEY: misc feature
LOCATION: (346860)..(346823)
OTHER INFORMATION: n is unidentified a, c, g, or t
NAME/KEY: misc feature
LOCATION: (317174)..(317193)
OTHER INFORMATION: n is unidentified a, c, g, or t
NAME/KEY: misc feature
LOCATION: (280353)..(280373)
OTHER INFORMATION: n is unidentified a, c, g, or t
NAME/KEY: misc feature
LOCATION: (271829)..(271848)
OTHER INFORMATION: n is unidentified a, c, g, or t
NAME/KEY: misc feature
LOCATION: (183872)..(183891)
OTHER INFORMATION: n is unidentified a, c, g, or t
NAME/KEY: misc feature
LOCATION: (170625)..(170645)
OTHER INFORMATION: n is unidentified a, c, g, or t
NAME/KEY: misc feature
LOCATION: (132680)..(132700)
OTHER INFORMATION: n is unidentified a, c, g, or t
NAME/KEY: misc feature
OTHER INFORMATION: n is a, c, g, or t
; US-09-771-208-20

Query Match 13.0%; Score 65.2; DB 9; Length 659158;
Best Local Similarity 73.8%; Pred. No. 5.2e-07;
Matches 96; Conservative 0; Mismatches 33; Indels 1; Gaps 1;

QY 317 GAGGCAAGAGAAATTTGCTGCTCAACAGAGCTTAGCTCCCTGGTGTGTTGTTCAAACTTTG 376
Db 14497 GAGGCAAGAGAAAGTCTGATTCACACAGACTCAG-TAGCTATGTATTTAAACTTTG 14439
QY 377 AAGTTTGACCAAGCAACTTATTTTGAATATTTTAAACAGAGCAACTTTGGGAAA 436
Db 14438 AAGTCTGACCTTGAGTATTTTATTTCAAGGATATTTAAGCATAGACACTTCAGTGAAA 14379
QY 437 AAGTTTCTT 446
Db 14378 AAAATTTCTT 14369

RESULT 4
US-10-175-523-57/c
; Sequence 57, Application US/10175523
; Publication No. US20030096264A1
; GENERAL INFORMATION:
; APPLICANT: Brockman, Jeffrey
; APPLICANT: Evans, David
; APPLICANT: Hook, Derek

```

; APPLICANT: Klimczak, Leszek
; APPLICANT: Laeng, Pascal
; APPLICANT: Palfreyman, Michael
; APPLICANT: Rajan, Pritih
; TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)
; FILE REFERENCE: 3235/11795-US3
; CURRENT APPLICATION NUMBER: US/10/175,523
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/299,151
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: US 60/317,828
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US 60/325,150
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/333,047
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: US 60/349,936
; PRIOR FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US 60/361,834
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 57
; LENGTH: 225883
; TYPE: DNA
; ORGANISM: Mus musculus
; US-10-175-523-57

```

Query Match 12.8%; Score 64.2; DB 14; Length 225883;

Best Local Similarity 61.7%; Pred. No. 5.8e-07; Matches 119; Conservative 0; Mismatches 73; Indels 1; Gaps 1;

```

QY 271 TGGGATATATCTGCAAGTGCATGATGTCAGAGGAGCAGCCATGTGAGCAAGAGAA 330
DB 215423 TGGAGGCTCTGTGACCTTCAAGCCCATCAAAAAGATGATCCCAAGAGCTAGAGAA 215364
QY 331 TTCTGCTCAACACAGCTTAGCTCCCTGCTTGTGCTTCAACTTTGAGTTGACACA 390
DB 215363 ATTGTTCACACACACTCA-ATGGCTGCTGTGATTTAAATTTCTGAGCTGACTCCA 215305
QY 391 AGCACTTATTTTGAATATTTAAACAGACACACTTTGGGAAAAGTTTCTTATGA 450
DB 215304 AGGAGTTTATTTTGGCAGGTTTAAAGCAGCCCAATTGTCTCAAACTTACTAGTGA 215245
QY 451 AAATTATCAAT 463
DB 215244 AAATTACTCACT 215232

```

RESULT 5

```

; US-10-087-192-1519
; Sequence 1519, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1519
; LENGTH: 48133
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature

```

```

; LOCATION: (1)...(48133)
; OTHER INFORMATION: n = A,T,C or G
; US-10-087-192-1519

```

Query Match 12.7%; Score 63.6; DB 13; Length 48133;

Best Local Similarity 67.9%; Pred. No. 3.9e-07; Matches 133; Conservative 0; Mismatches 59; Indels 4; Gaps 3;

```

QY 268 AGTGGGATATCTGCAAGTGCATGATGTCAGAGGAGCAGCCATGTGAGCAAGAG 327
DB 31444 ATGTAGGAGCTCTGTACACACACAGCAGCGCTTAAGCC-ACATGAGTAAGAG 31502
QY 328 GAATCTGCTCAACACAGCTTAGCTCCCTGCTGTGTTCAACTTTGAGAGTTGACC 387
DB 31503 GAAGTCTGCTCAACACACTTAGTACC-GGTGTTGTTTAACTTCAAGGTTCAACA 31561
QY 388 ACAAGCACTTATTTTGAATATTTAAACAGACACACTTTGGGAAAAGTTTCTTA 447
DB 31562 CTGGCAATTTATGTAGTATTTAAGCACAACACAA--TCTGAAAATAGTTTCTTG 31619
QY 448 TGAATATATCAAT 463
DB 31620 TGATAGTTAATCAAT 31635

```

RESULT 6

```

; US-10-085-117-121/C
; Sequence 121, Application US/10085117
; Publication No. US2003023334A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: 529452000121
; CURRENT APPLICATION NUMBER: US/10/085,117
; CURRENT FILING DATE: 2002-02-27
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 121
; LENGTH: 47884
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: variation
; LOCATION: (1)...(47884)
; OTHER INFORMATION: n = any nucleotide
; US-10-085-117-121

```

Query Match 12.5%; Score 62.4; DB 17; Length 47884;

Best Local Similarity 57.5%; Pred. No. 8.8e-07; Matches 131; Conservative 0; Mismatches 96; Indels 1; Gaps 1;

```

QY 270 GTGGGATATCTGCAAGTGCATGATGTCAGAGGAGCAGCCATGTGAGCAAGAG 329
DB 40174 GTGGGATCTCTGAAACACTTAAACAGAGGCTGTGATTCAGAGAGAGAGAGAA 40115
QY 330 ATTCTGCTCAACACAGCTTAGCTCCCTGCTGTGTTCAACTTTGAGAGTTGACCAC 389
DB 40114 AGTCTGTCTCAAGTACTCAGTGACAGTTTTT-GTCCAAACTTGGAGTACGACC 40056
QY 390 AAGCACTTATTTTGAATATTTAAACAGACACACTTTGGGAAAAGTTTCTTATG 449
DB 40055 AAGTAGTTTATTTTGAAGCAGCACAATTTCAAAATGTTCTGTGATTAATATCACT 39996
QY 450 AAATTATCAATTAAGCTTAAGGATGATGATCAATTAATGCTTT 497
DB 39995 CCTGTGTGACAGAAAGCTTAGGAGATGCTACTGAACCTTTGT 39948

```

RESULT 7

US-10-417-375-15

Query Match: 61.6%; Score 61.6; DB 18; Length 145068;
 Best Local Similarity 58.3%; Pred. No. 2.7e-06;
 Matches 126; Conservative 0; Mismatches 89; Indels 1; Gaps 1;

QY 275 GATATTCGCAAGTCCATGATGTCAGAGGACAGCCATGTGACGCATAGAGATTCT 334

Db 101150 GGCTCTGTGCAGACCAAGCCAGCCATGTAAAGATTCACAGAGGCAAGAAAGCGCATTTG 101209

Query Match	12.3%	Score 61.4	DB 11	Length 67832
Best Local Similarity	77.9%	Pred. No. 2.1e-06		
Matches 74	Conservative 0	Mismatches 21	Indels 0	Gaps 0
QY	407	CATATTAAACGAGACACAACCTTGGAAAAAGTTTCTATGAAATATACAAATAA	466	

Qy	Db
467	62076
GCTTAAGCATGACTACATTAAATGCTTTGCA	CACATTTAAACACGTAAAGACTTTGGGGAAAAATTCTTCTGTACAGTATACGGCAAA
501	62017
62016	62016
GCTTTAGGCATGGCTACATCAAAACTTTTGTAA	61982

```

RESULT 10
US-10-087-192-289
? Sequence 289, Application US/10087192
? Publication No. US20020182586A1
? GENERAL INFORMATION:
? APPLICANT: Morris, David W.
? APPLICANT: Engelhard, Eric K.
? TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
? TITLE OF INVENTION: CANCER
? FILE REFERENCE: 52945200122
? CURRENT APPLICATION NUMBER: US/10/087,192
? CURRENT FILING DATE: 2002-03-01
? PRIOR APPLICATION NUMBER: US 09/747,377
? PRIOR FILING DATE: 2000-12-22
? PRIOR APPLICATION NUMBER: US 09/798,586
? PRIOR FILING DATE: 2001-03-02
? NUMBER OF SEQ ID NOS: 2059
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 289
? LENGTH: 210528
? TYPE: DNA
? ORGANISM: Mus musculus
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (1)...(210528)
? OTHER INFORMATION: n = A,T,C or G
US-10-087-192-289

```

	Query Match	Similarity	12.1%	Score 60.8	DB 13	length 210528
	Best Local	Similarity 63.4%	Pred. No.5.8e-06			
	Matches 109	Conservative 0	Mismatches 62	Indels 1	Gaps 1	
Qy	284	AAAGTCGCATGAGTGCAGAGGGAACGCCAATGTGAGGCAAGAGAAATTCGTGCTCAACA	343			
Db	172729	AAATCTCCCTGGGTAGCAGAGTAACCTCTACAAAGAAAGCAAGTTCCTTCAACC	172788			
Qy	344	CAGCTTAGCTCCCTGTGTGTGTTCAAACCTTGAGAGTTGATCCACAGCACTTTATTTT	403			
Db	172789	CAACTCA-CTAGTCAGTGTGTGTTAGCCTTCGTGGGATCTGACCCTTAGAGTTTACTAT	172847			
Qy	404	TGACATATTAAACAGAGCAACCTTGGGAAAAGTTTCTATGAAATTT	455			
Db	172848	AGGCACATTTAAGTACAGTACAAATTTGATGAACATGCATTTCTGTGCTAAAT	172899			

```

RESULT 11
US-10-322-281-587/c
: Sequence 587, Application US/10322281
: Publication No. US20040126762A1
: GENERAL INFORMATION:
: APPLICANT: David W. Morris
: APPLICANT: Marc S. Malandro
: TITLE OF INVENTION: Novel Compositions and Methods in Cancer
: FILE REFERENCE: 529452001000
: CURRENT APPLICATION NUMBER: US/10/322,281
: CURRENT FILING DATE: 2002-12-17
: NUMBER OF SEQ ID NOS: 866
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 587
: LENGTH: 121129
: TYPE: DNA
: ORGANISM: Mus musculus
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)...(121129)

```

PRIOR APPLICATION NUMBER: US 09/798,586
PRIOR FILING DATE: 2001-03-02
NUMBER OF SEQ ID NOS: 301
SOFTWARE: PatentIn version 3.1
SEQ ID NO 199
LENGTH: 96599
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: misc.feature
LOCATION: (4185)..(4208)
OTHER INFORMATION: "n" at positions 4189 through 4208 can be any base.
FEATURE:
NAME/KEY: misc.feature
LOCATION: (9652)..(10001)
OTHER INFORMATION: "n" at positions 9652 through 10001 can be any base.
FEATURE:
NAME/KEY: misc.feature
LOCATION: (27805)..(27824)
OTHER INFORMATION: "n" at positions 27805 through 27824 can be any base.
FEATURE:
NAME/KEY: misc.feature
LOCATION: (54062)..(54139)
OTHER INFORMATION: "n" at positions 54062 through 54139 can be any base.
FEATURE:
NAME/KEY: misc.feature
LOCATION: (88386)..(88405)
OTHER INFORMATION: "n" at positions 88386 through 88405 can be any base.
US-09-997-722-199

Query Match 11.7%; Score 58.6; DB 11; Length 96599;
Best Local Similarity 77.1%; Pred. No. 1.7e-05;
Matches 84; Conservative 0; Mismatches 24; Indels 1; Gaps 1;

QY 387 CACAGCACTTATTTTGGACATTTTAACAGAGCACTTGGAAAAAGTTTCTT 446
DB 46014 CTCGAACGTTTCATTTTGAACATTTTAACAGAGCACTTGGAAAAAGTTTCTG 46072
QY 447 ATGAATTTATCATCAATTAAGCTTAAGCATGACTTAAATGCTT 495
DB 46073 GTGACAGTTATCATCAAGTAAGCTTAAGCATGACTTAAATGCTT 46121

RESULT 14
US-10-085-117-235
Sequence 235, Application US/10085117
Publication No. US2003023233A1
GENERAL INFORMATION:
APPLICANT: Morris, David W.
APPLICANT: Engelhard, Eric K.
TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
FILE REFERENCE: 529452000121
CURRENT APPLICATION NUMBER: US/10/085,117
PRIOR FILING DATE: 2002-02-27
PRIOR APPLICATION NUMBER: US 09/798,586
NUMBER OF SEQ ID NOS: 361
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 235
LENGTH: 33488
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: variation
LOCATION: (1)...(33488)
OTHER INFORMATION: n = any nucleotide
US-10-085-117-235

Query Match 11.6%; Score 58; DB 17; Length 33488;
Best Local Similarity 66.1%; Pred. No. 1.5e-05;
Matches 115; Conservative 0; Mismatches 55; Indels 4; Gaps 2;
QY 293 GAGTGGAGAGGACAGCAATGTGAGGCAAGAAAGAAATTTGGCTCAACACAGCT--T 349

DB 5303 GATGCCACAGAAACATGCAAGGGGACAAAAGTGTTTCATTTTAACACACTAAGT 5362
QY 350 AGCTCCCTGATGTTGGTTTAACTTTGAGAGTTTGAACACAGCACTTATTTTGGACAT 409
DB 5363 AAGTACCGAGTATGTGTTCAAACTTTGAGAGTGTGACCCCAAGTATTTATTTAGGCTT 5422
QY 410 ATTTAAACAGAGCACTTTGGAAAAAGTTTCTTATGAATAATTCACAAT 463
DB 5423 ATTTAAGCAGACAGCACAC-TTGACAGACATTTCTGTGAATATTAATTAAGT 5475

RESULT 15
US-10-087-192-1237
Sequence 1237, Application US/10087192
Publication No. US20020182586A1
GENERAL INFORMATION:
APPLICANT: Morris, David W.
APPLICANT: Engelhard, Eric K.
TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
FILE REFERENCE: 529452000122
CURRENT APPLICATION NUMBER: US/10/087,192
PRIOR FILING DATE: 2002-03-01
PRIOR APPLICATION NUMBER: US 09/747,377
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/798,586
NUMBER OF SEQ ID NOS: 2059
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1237
LENGTH: 75101
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)...(75101)
OTHER INFORMATION: n = A,T,C or G
US-10-087-192-1237

Query Match 11.6%; Score 58; DB 13; Length 75101;
Best Local Similarity 57.3%; Pred. No. 2.3e-05;
Matches 126; Conservative 0; Mismatches 90; Indels 4; Gaps 1;

QY 219 TATATTGCAATTTATGAGCAAGCTAAATGTACTCGGGTATGATGACTCCAGTGGGATA 278
DB 35225 TTTAATCCAAAATTCAGAAACAAATGAGGGCTAAGATGGGAATGCTCAGTGTGGTT 35284
QY 279 TCTGCAAGTCCATGAGTGGGAGAGGACAGCCATGTGAGGCAAGAAATTCGTGCT 338
DB 35285 CCGCAAGCTTGAAGGCTGGAAAGAAACATCTAGTGAAGTGAAGAAAGCTTGCTT 35344
QY 339 CACACAGCTTAGCTCCCTGCTGTGTGTTCAAACTTTGAGAGTTTGAACACAGCACTT 398
DB 35345 CAGCCGACCTTGACA-----GATGCCGGGCCACACTTTGAAGTCTGACACTGGGTGTTT 35400
QY 399 ATTTTGAATATTTTAAACAGAGCACTTTGGAAAAA 438
DB 35401 ATTTTGAATATTTTAAACAGAGCACTTTTGAAGAAA 35440

Search completed: February 10, 2005, 18:16:57
Job time : 372.063 secs

RESULT 2	732 bp	DNA	linear	GSS	05-MAR-2001
LOCUS	AZ900293				
DEFINITION	RPCT-24-212G17.TU RPCT-24 Mus musculus genomic clone				
ACCESSION	RPCT-24-212G17				
VERSION	AZ900293				
KEYWORDS	AZ900293.1 GI:13219238				
SOURCE	GSS.				
ORGANISM	Mus musculus (house mouse)				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	1 (baes 1 to 732)				
TITLE	Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akintret,B., Levins,M., Tsegay,G., Geer,K., Kroll,M., Shvartsbeyn,A., Gebregeorgis,E., Russell,D., de Jong,P. and Fraser,C.M.				
JOURNAL	Mouse BAC End Sequences from Library RPCT-24				
COMMENT	Unpublished (1999)				
	Contact: Shaying Zhao				

Clones: 5
Clones are derived from the mouse BAC library RPcl-24. For BAC library availability, please contact Plateri de Jong (pdjong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end sequence: http://www.tigr.org/cdb/bac_ends/mouse/bac_end_intro.html
Plate: 212 row: G column: 17
Seq primer: 586
Class: BAC ends

```

location/Qualifiers
1..732
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPC1-24-212G17"
/sex="Male"
/cell_type="Spleen/Brain"
/clone_id="RPC1-24"
/notice="Vector: pTRABAC1, Site_1: BamH1, Site_2: BamH1;
RPC1-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTRABAC1 cloning vector at the
BamH1 sites using MboI partially digested male C57BL/6J
DNA."

```

	Query Match	Similarity	75.0%	Score	118.4	DB 8	Length	732
	Best Local	Similarity	75.0%	Pred	No. 1.7e-21			
	Matches	174	Conservative	0	Mismatches	56	Indels	2
								Gaps
QY	270	GTGGGATATCTGCAGAGTCCATGAGTGCGCAGAGGAGACGCCAATGTGAGGCAGAGA	329					
Db	55	GTGGGGGTCCCTCAACAACTATTATGACTCTCAGAGAGATGATGCTCATAGGCAAGAAA	114					
QY	330	ATTTCGGCTCAACACAGCTTAGCTCCCGTGSTGTGTTCAAACTTGAAGTTGACAC	389					
Db	115	AATTCAGTTCAACAAT-TTAACTAGCTGGGTGGGTCAAACTTGAAGTTGACACCC	173					
QY	390	AAGCACTTATTTTGGACATATTTTAAACAGACACACTTTGGAAAAAGTTTCTTATG	449					
Db	174	GAGTAGTTTATTTTGA-TTTCTTAAGCAATAGTAAACAACCTTGGAAAAAGTTTCTTATG	232					
QY	450	AAATATATCACAATAAAGCTTAAGCAGTACATCAATTAAAGCCTTTGCAA	501					
Db	233	ACAAATATTATCAATAAAGTTTAGACATGACTACATTAATACTTCCTTTGCAA	284					

[illegible]

Email: szabo@tigr.org
Clones are derived from the mouse BAC library RPCT-23. For BAC library availability, please contact Peter de Jong (pjeteredjong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.html>) or from Resea ch Genetics (info@resgen.com). BAC end pase: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 216 row: E column: 16
Seq primer: SP6
Class: BAC ends.

```

Location/Qualifiers
1. 470
   /organism="Mus musculus"
   /mol_type="genomic DNA"
   /strain="C57BL/6J"
   /db_xref="taxon:10090"
   /clone="RPci-23-216E16"
   /sex="Female"
   /lab_host="DH10B"
   /clone_11b="RPci-23"
   /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
   EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
   brain genomic DNA was isolated and partially digested
   with a combination of EcoRI and EcoRI Methylase. Size
   selected DNA was cloned into the pBACe3.6 vector at the
   EcoRI sites. The ligation products were transformed into
   DH10B electrocompetent cells (BRL Life Technologies)."

```

ORIGIN	Query Match	Best Local Similarity	19.8%;	Score 99.2;	DB 8;	Length 470;
	Query Match	Best Local Similarity	69.7%;	Pred. NO. 3e-16;		
	Matches 163;	Conservative	0;	Mismatches 68;	Indels 3;	Gaps 2
QY	270	GTGGGATATCTGCAGATGCCATGATGGCAGAGGACACCCAAATCTGAGCGAAGAAGA	329			
Db	360	GTGAAGGCTCTCGCAACAGCTATATGGCAGCAGAGAAACAGTCACAGGGAGGCAAGGANA	301			
QY	330	AATTCGGCTCAACACAGCTTAGCTCCGATGGTGGTCAAACTTTGAGATTGACAC	389			
Db	300	AATCTGGTTCAGTACAAATGTAAGTACTGAGAGCTGGTTCACCTTTTAGAGTGTGACACT	241			
QY	330	AAGCACTTATTTTGGACATAATTTAAACAGACACAACCTTTGGAAAAAGTTTCTTAG	449			
Db	240	GGGCAATTTGGTTTGGACACATTTAAGTTCAGTAC-AGTTTTGGAAAAATATTTCTGCTG	182			
QY	450	AAATATATCACAAATAAG--CTTAGGAGATGACTCATTTAAATGGCTTTGGCA	501			
Db	181	ACAATATATCACAAAAAGGCTTTAAAGCAATGGCTCATCATTAACATCTTTTGGCA	128			

[illegible]

	AZ413362/c	532 bp	DNA	linear	GSS 03-Oct-2000
LOCUS					
DEFINITION	c1m0197G04F Mouse 10kb plasmid UNGC1M library Mus musculus genomic				
ACCESSION	AZ413362				
VERSION	1				
KEYWORDS	GI:10537375				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.				
AUTHORS	1 (bases 1 to 532) Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Irlam,H., Longacre,S., Mahmood,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A. and Wright,D.,Weiss,R. Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts Unpublished (2000) Contact: Robert B. Weiss University of Utah Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA Tel.: 801 585 5606 Fax: 801 585 7177 Email: dunn@genetics.utah.edu Insert Length: 10000 Std Error: 0.00 Plate: 0197 row: G column: 04 Seq primer: CGTTTAAACGACGCCACAGT Class: Plasmid ends High quality sequence stop: 532.				
FEATURES					
source	Location/Qualifiers 1..532 /organism="Mus musculus" /mol_type="genomic DNA" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="UNGCM0197G04" /sex="Male" /lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-" /clone_id="Mouse 10kb plasmid UNGC1M library" /note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 [gi 4732114 gb AF129072.1], a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adaptor-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."				
ORIGIN					
Query Match	18.1%; Score 90.8; DB 8; Length 532;				
Best Local Similarity	66.7%; Pred. No. 6.5e-14;				
Matches 160; Conservative	0; Mismatches 77; Indels 3; Gaps 2				
OY	262 GACTCCAGGTGGGATATCTGCAGATGCCATGACGACGAGGACAGCCATGTGAGGC 321				
DB	513 GTCTTTTAGAGGGGGTTCCTCAACAATTCTCGTGTTGACAGAAAACAATTGGACAAGGT 454				
OY	322 AAGAAGGAATTCGTGCTCAACACAGCTTAGCTCCTCGTGTTGGTTCAAACCTTGAGAGT 381				

Db 453 AAGACAGAACCTGTTCAGTACACTTGTGGCCTG--GCTCATTCAAACCTTCGAGT 396

OY 382 TTGACCACAGACACTTTATTTTGGACATATTTAAACAGACAACTTTGGAAAAAGTT 441

Db 395 CTGAGCCCAATGATTTATTTTGGACATATTTAAATAGTACAACTTTGGG-AAAGTTT 337

OY 442 TTTCTATGAAATTTATCACAATTAAGCTTAAGGCATGACTATTAATAATGCTTTGGAA 501

Db 336 TCCCTGTGACATATATCAACAATTAAGCTTCGGGCAATGGCTGCATTAAACCTCTTGCA 277

RESULT 6
A2555307 750 bp DNA linear GSS 20-NOV-2000
LOCUS RPCI-23-175N14.TU RPCI-23 Mus musculus genomic clone
DEFINITION RPCI-23-175N14, genomic survey sequence.
ACCESSION A2555307
VERSION A2555307.1 GI:11235127
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 750)
Zhao, S., Niernann, W., Feldblyum, T., Malek, J., Shatsman, S., Akinter, B., Levins, M., McGarr, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P. and Fraser, C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@igr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea.ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac/ends/mouse/bac_end_intro.html
Plate: 175 row: N column: 14
Seq primer: SP6
Class: BAC ends.

FEATURES
source
1..750
Location/Qualifiers
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-175N14"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: PBAC3.6; Site 1: EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the PBAC3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

ORIGIN
Query Match 17.1%; Score 85.6; DB 8; Length 750;
Best Local Similarity 66.5%; Pred. No. 2e-12;
Matches 157; Conservative 0; Mismatches 69; Indels 10; Gaps 2;

OY 270 GTGGGATATCTGCAAGTGCATGAGCGCAGAGGACACCCATGTGAGGCAAGAAGA 329

Db 253 GTGAGAGTCCGCCCCACACACTTAAGGCTTACTGGGAGCAAAACGAAAACAAAGGA 312

OY 330 ATTCTGCTCAACACAGCTTAGCTCCCTGGTGTGTTCAAACTTTGAGAGTTGACAC 389

Db 313 AATCTGCTAGTGAATTAAGCTGGCGG-----TTCAACTTCGGCGTTGACTCT 366

OY 390 AAGACATTTATTTTGAATTTAAACA-----GAGACAACTTTGGAAAAAATTTCT 445

Db 367 AAACAGTTTACTTTATTTATTTTGAATTTTGAACAACATCTGGAAAAAGTTTCT 426

OY 446 TATGAATTTATCAATTAAGCTTAAGGCATGACTACATTAATAATGCTTTGCA 501

Db 427 GATGCAATTAACAAATTAAGCTTAAGGCATGCTGACTGATTAACATCTTTACA 482

RESULT 7
CB770238 409 bp mRNA linear EST 16-MAY-2003
LOCUS AMANNUD:SRCP2-00001-E2-A srp2 (10233) Rattus norvegicus cDNA clone
DEFINITION srp2-00001-e2 5', mRNA sequence.
ACCESSION CB770238
VERSION CB770238.1 GI:29858629
KEYWORDS EST.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 409)
Amgen EST Program.
Amgen Rat EST Program
Unpublished (2003)
Contact: Dan Fitzpatrick
Amgen, Inc
One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA
Tel: 805 447-4881
Plate: 00001 row: e column: 2.

FEATURES
source
1..409
Location/Qualifiers
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="srp2-00001-e2"
/issue_type="choroid plexus brain"
/clone_lib="srp2 (10233)"
/note="Vector: pSPOR1; Site_1: SalI; Site_2: NotI; choroid plexus brain region"

ORIGIN
Query Match 16.9%; Score 84.8; DB 6; Length 409;
Best Local Similarity 71.1%; Pred. No. 2.8e-12;
Matches 128; Conservative 0; Mismatches 47; Indels 5; Gaps 1;

OY 317 GAGCAAGAAGAAATTTGCTCAACACAGCTTACCTCCGTGTGTTGTTCAACTTGG 376

Db 43 GAGTTAATAGAAATCTTGTTCAATTAAGCTTACT-----GGCCCGCTTCAAGCTCT 97

OY 377 AAGATTGACCAACAAGCACTTATTTTGGACATATTTAAACGAGACACAACCTTTGGAAA 436

Db 98 GGAGCTGACACCAAGCAGTCTATTTTGGACATATTTAAACAGCTAAACCTTTGGGGA 157

OY 437 AAGTTTCTTAAGAAATTAATCACAATTAAGCTTAAGGCATGACTACATTAATGCTTT 496

Db 158 AAGTTTCTTGTGATTAATTAACAACAAGCTCCAGGTCTGCTGCTTGAATCTCAT 217

RESULT 8
B2189040/c 500 bp DNA linear GSS 11-OCT-2002
LOCUS CH230-378B15.TU CHORI-230 Segment 2 Rattus norvegicus genomic clone
DEFINITION CH230-378B15, genomic survey sequence.
ACCESSION B2189040
VERSION B2189040.1 GI:22843505
KEYWORDS GSS.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus

REFERENCE
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 500)

Shvartsbeyn, A., Gebregorgis, E., Overton, L., Russell, D., Chen, D., Riggs, P., de Jong, P. and Fraser, C.M.

Rat BAC End Sequences from Library CHORI-230 MboI segment

Unpublished (1999)

TITLE
JOURNAL

Other GSSs: CH230-378B15.TV

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the rat BAC library CHORI-230

(<http://www.chori.org/bacpac/rat230.htm>). For BAC library

availability, please contact Pieter de Jong (pdejong@mail.cho.org).

Clones may be purchased from BACPAC Resources

(<http://www.chori.org/bacpac/orering/information.htm>). BAC end

page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html

Plate: 378 row: B column: 15

Seq primer: Sp6

Class: BAC ends.

Location/Qualifiers

1..500

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/strain="BN/SsNHsd/MCW"

/db_xref="taxon:10116"

/clone="CH230-378B15"

/sex="Female"

/cell_type="Brain"

/clone_1lib="CHORI-230 Segment 2"

/note="Vector: PTARBAC1.3; Site_1: MboI; Site_2: MboI;

CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by

Pieter de Jong"

ORIGIN

Query Match 16.6%; Score 83.4; DB 8; Length 500;

Best Local Similarity 75.8%; Pred. No. 7.1e-12;

Matches 116; Conservative 0; Mismatches 36; Indels 1; Gaps 1;

349 TAGCTCCCTGGTGGTGAACCTTGAAGTGAACCAAGACCTTATTTTGA 408

283 TTGTAGCTGGTCTAGTCAAGTCCGAGTCTGAC -CAACAGTTATTTTGA 225

409 TATTAAACAGACCAACTTTGGAAAAAGTTTCTTATGAAAAATATCAATAAAGC 468

224 TCGTTAAACATAGCAACCTTGAATAAAAGTCTCTTGTGCAATTACATGATAAGC 165

469 TTAAGGCACTACATTAATAATGCTTTGCA 501

164 TTAAGGATGACCCCATGAAACTCTTTGGCA 132

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

TITLE
JOURNAL

Shvartsbeyn, A., Gebregorgis, E., Overton, L., Russell, D., Chen, D., Riggs, P., de Jong, P. and Fraser, C.M.

Rat BAC End Sequences from Library CHORI-230 MboI segment

Unpublished (1999)

Other GSSs: CH230-334G8.TV

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the rat BAC library CHORI-230

(<http://www.chori.org/bacpac/rat230.htm>). For BAC library

availability, please contact Pieter de Jong (pdejong@mail.cho.org).

Clones may be purchased from BACPAC Resources

(<http://www.chori.org/bacpac/orering/information.htm>). BAC end

page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html

Plate: 334 row: G column: 8

Seq primer: Sp6

Class: BAC ends.

Location/Qualifiers

1..812

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/strain="BN/SsNHsd/MCW"

/db_xref="taxon:10116"

/clone="CH230-334G8"

/sex="Female"

/cell_type="Brain"

/clone_1lib="CHORI-230 Segment 2"

/note="Vector: PTARBAC1.3; Site_1: MboI; Site_2: MboI;

CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by

Pieter de Jong"

ORIGIN

Query Match 16.5%; Score 82.8; DB 8; Length 812;

Best Local Similarity 67.5%; Pred. No. 1.2e-11;

Matches 131; Conservative 0; Mismatches 62; Indels 1; Gaps 1;

270 GTGGGATATCTTCAAGTCCATGAGTGGAGGAGCAACCAATGTAGGCAAGCA 329

67 GGGGTGCTCTGAGGCACACAGTCAGATCTGTGGAGGAGAAAGGT 126

330 ATTCTGCTCAACACGCTTACCTCTGTGTGTTCAACTTGAAGAGTTTGA 389

127 AGTTTGCTTAAACA-AGACTATATAGCAGTGTGTTCAAACTTGAAGAGTTGACCC 185

390 AAGCACTTATTTTGAATATTTAAACAGACCAACTTGGAAAAAGTTTCTTATG 449

186 AAGTAGTTTATGTAGGCAATTTCAACACAAACAAATTTGGTAAAAATTTTCTGGAG 245

450 AAAATTATCAAT 463

246 ATATATGATCAAT 259

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

TITLE
JOURNAL

Shvartsbeyn, A., Gebregorgis, E., Overton, L., Russell, D., Chen, D., Riggs, P., de Jong, P. and Fraser, C.M.

Rat BAC End Sequences from Library CHORI-230 MboI segment

Unpublished (1999)

Other GSSs: CH230-334G8.TV

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the rat BAC library CHORI-230

(<http://www.chori.org/bacpac/rat230.htm>). For BAC library

availability, please contact Pieter de Jong (pdejong@mail.cho.org).

Clones may be purchased from BACPAC Resources

(<http://www.chori.org/bacpac/orering/information.htm>). BAC end

page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html

Plate: 334 row: G column: 8

Seq primer: Sp6

Class: BAC ends.

Location/Qualifiers

1..812

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/strain="BN/SsNHsd/MCW"

/db_xref="taxon:10116"

/clone="CH230-334G8"

/sex="Female"

/cell_type="Brain"

/clone_1lib="CHORI-230 Segment 2"

/note="Vector: PTARBAC1.3; Site_1: MboI; Site_2: MboI;

CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by

Pieter de Jong"

ORIGIN

Query Match 16.5%; Score 82.8; DB 8; Length 812;

Best Local Similarity 67.5%; Pred. No. 1.2e-11;

Matches 131; Conservative 0; Mismatches 62; Indels 1; Gaps 1;

270 GTGGGATATCTTCAAGTCCATGAGTGGAGGAGCAACCAATGTAGGCAAGCA 329

67 GGGGTGCTCTGAGGCACACAGTCAGATCTGTGGAGGAGAAAGGT 126

330 ATTCTGCTCAACACGCTTACCTCTGTGTGTTCAACTTGAAGAGTTTGA 389

127 AGTTTGCTTAAACA-AGACTATATAGCAGTGTGTTCAAACTTGAAGAGTTGACCC 185

390 AAGCACTTATTTTGAATATTTAAACAGACCAACTTGGAAAAAGTTTCTTATG 449

186 AAGTAGTTTATGTAGGCAATTTCAACACAAACAAATTTGGTAAAAATTTTCTGGAG 245

450 AAAATTATCAAT 463

246 ATATATGATCAAT 259

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

TITLE
JOURNAL

Shvartsbeyn, A., Gebregorgis, E., Overton, L., Russell, D., Chen, D., Riggs, P., de Jong, P. and Fraser, C.M.

Rat BAC End Sequences from Library CHORI-230 MboI segment

Unpublished (1999)

Other GSSs: CH230-334G8.TV

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the rat BAC library CHORI-230

(<http://www.chori.org/bacpac/rat230.htm>). For BAC library

availability, please contact Pieter de Jong (pdejong@mail.cho.org).

Clones may be purchased from BACPAC Resources

(<http://www.chori.org/bacpac/orering/information.htm>). BAC end

page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html

Plate: 334 row: G column: 8

Seq primer: Sp6

Class: BAC ends.

Location/Qualifiers

1..812

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/strain="BN/SsNHsd/MCW"

/db_xref="taxon:10116"

/clone="CH230-334G8"

/sex="Female"

/cell_type="Brain"

/clone_1lib="CHORI-230 Segment 2"

/note="Vector: PTARBAC1.3; Site_1: MboI; Site_2: MboI;

CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by

Pieter de Jong"

ORIGIN

Query Match 16.5%; Score 82.8; DB 8; Length 812;

Best Local Similarity 67.5%; Pred. No. 1.2e-11;

Matches 131; Conservative 0; Mismatches 62; Indels 1; Gaps 1;

270 GTGGGATATCTTCAAGTCCATGAGTGGAGGAGCAACCAATGTAGGCAAGCA 329

67 GGGGTGCTCTGAGGCACACAGTCAGATCTGTGGAGGAGAAAGGT 126

330 ATTCTGCTCAACACGCTTACCTCTGTGTGTTCAACTTGAAGAGTTTGA 389

127 AGTTTGCTTAAACA-AGACTATATAGCAGTGTGTTCAAACTTGAAGAGTTGACCC 185

TITLE Direct Submission
JOURNAL Submitted (20-FEB-2004) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. <http://www.sanger.ac.uk/MICER>
FEATURES
source
1. 871
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="MHPN417h21"
/clone_lib="MHPN"
ORIGIN
Query Match 16.5%; Score 82.6; DB 9; Length 871;
Best Local Similarity 61.7%; Pred. No. 1.4e-11;
Matches 171; Conservative 0; Mismatches 94; Indels 12; Gaps 2;
232 TGAGCAGCTAAATGTAATCTCGGAGTACATGCTCAGGTGGGATATCTGCAAGTGC 291
DB 722 TCACACAGCTTACACTGACCTTACACCCCGATCTACTAGGGCTCTGCAATACCA 663
QY 292 TGAGTGCAGAGGACACGCCAATGTGAGCAAGAAAGAAATCTGCTCAACAGCTTGA 351
DB 662 TGCTTACCTGATTTGTTATTCATGTGAGCAAGAAAGTCTGTTCAACAGCACTGAG 603
QY 352 CTCCTGTGTGTGTTCAAACTTTGAGAGTTGACCAAGCAACTTATTTTGCATAT 411
DB 602 AGGCC-GGTGTGTGTCCAACTTTGAGAGTCCAAATGCTGGGTGTTATTTTGAAGCACAT 544
QY 412 TTAACAGAGCAACACTTTGGGAAAAAGTTTCTTATGAAAA-----TTATCAC 460
DB 543 TTAGACAGCAGCAATTTGTAAGAAAAAGATTTCTGTGATTAACGATCCTGTGACAC 484
QY 461 AATAAGCTTAAGCAGTACATCAATTAATGCTTT 497
DB 483 AACAAATCATAGATATGATTAATCAATGAATCTTAT 447
RESULT 11 950 bp DNA linear GSS 06-JUL-2004
LOCUS CR199109/c
DEFINITION Forward strand read from insert in 5'HPRT insertion targeting and
chromosome engineering clone MHPN34906, genomic survey sequence.
ACCESSION CR199109
VERSION CR199109.1 GI:49977958
KEYWORDS GSS; genome survey sequence; MICR.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 950)
AUTHORS Adams,D.J., Biggs,F.J., Cox,A.V., Davies,R.M., van der Weyden,L.,
Jonkers,J., Smith,J., Plumb,R.W., Taylor,R.G., Nishijima,T., Yu,Y.,
Rogers,J. and Bradley,A.
TITLE Direct Submission
JOURNAL Submitted (20-FEB-2004) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. <http://www.sanger.ac.uk/MICR>
FEATURES
source
1. 950
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="MHPN34906"
/clone_lib="MHPN"
ORIGIN
Query Match 16.5%; Score 82.6; DB 9; Length 950;
Best Local Similarity 61.7%; Pred. No. 1.4e-11;
Matches 171; Conservative 0; Mismatches 94; Indels 12; Gaps 2;
232 TGAGCAGCTAAATGTAATCTCGGAGTACATGCTCAGGTGGGATATCTGCAAGTGC 291
DB 725 TCACACAGCTTACACTGACCTTACACCCCGATCTACTAGGGGTCTGCAATACCA 666

QY 292 TGAGTGCAGAGGACACGCCAATGTGAGCAAGAAATTTCTGCTCAACAGCTTAG 351
DB 665 TGCTTACCTGATTTGTTATTCATGTGAGCAAGAAAGTCTGTTCAACAGCACTGAG 606
QY 352 CTCCTGTGTGTGTTCAAACTTTGAGAGTTGACCAAGCAACTTATTTTGCATAT 411
DB 605 AGGCC-GGTGTGTGTCCAACTTTGAGAGTCCAAATGCTGGGTGTTATTTTGAAGCACAT 547
QY 412 TTAACAGAGCAACACTTTGGGAAAAAGTTTCTTATGAAAA-----TTATCAC 460
DB 546 TTAAGCAGCAGCAATTTGTAAGAAAAAGATTTCTGTGATTAACGATCCTGTGACAC 487
QY 461 AATAAGCTTAAGCAGTACATCAATTAATGCTTT 497
DB 486 AACAAATCATAGATATGATTAATCAATGAATCTTAT 450
RESULT 12 714 bp DNA linear GSS 04-JUN-2004
LOCUS AG489174
DEFINITION Mus musculus molossinus DNA, clone:MSMg01-386C04.TJ, genomic survey
sequence.
ACCESSION AG489174
VERSION AG489174.1 GI:48196404
KEYWORDS GSS.
SOURCE Mus musculus molossinus
ORGANISM Mus musculus molossinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE BAC end Sequences of Library MSMg01
JOURNAL Unpublished
2 (bases 1 to 714)
Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
AUTHORS Direct Submission
TITLE Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
1-7-22 Suehiro-cho,Tsukuba-shi, Ibaraki, Japan
JOURNAL (E-mail:hattori@gsc.riken.jp, URL:<http://hgp.gsc.riken.go.jp/>,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the mouse BAC library MSMg01. For BAC
library availability, please contact Kunihisa Abe (abe@rtc.riken.jp).
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koydai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : TJ
LIBRARY : BAC63.6
Vector : EcoRI
R.Site 1 : EcoRI
R.Site 2 : EcoRI
FEATURES
source
1. 714
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-386C04.TJ"
/sex="male"
/tissue type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN
Query Match 16.4%; Score 82; DB 9; Length 714;
Best Local Similarity 69.2%; Pred. No. 1.9e-11;
Matches 144; Conservative 0; Mismatches 55; Indels 9; Gaps 2;
298 GCAGAGGACAGCCCAATGTGAGCAAGAAATTTCTGCTCAACAGCTTAGCTCCCT 357
DB 371 GCTGGAGACAAACAGAAAGACAGAAAGAAATTCGCTACAGCAACTTAGCT---- 426

358 GGTTGTTGTTCAAACTTTGAGCTTGACCAAGACACTTATTTTGGACATATTTAAAC 417
427 -GGGCGGGTTCAAACTTTGCGCTTGACCTTAAAGATTTACTTATTTATTTTGGACA 485
418 A-----GAGCACAACCTTTGGAAAAGTTTCTATGAAAATTTCAATTAAGCTTAAG 473
486 TATTTAAAGCAACTCTGGGAAAAGCTTCTATTAACATTAACATTAAGGTTTAAAG 545
474 GCATGACTACATTAATAATGCTTTGGCA 501
546 GCATGCTACTTGAATCACTCTTACTA 573

RESULT 13
A2776986 455 bp DNA linear GSS 16-FEB-2001
LOCUS 2M0011N06F Mouse 10kb plasmid UGCG1M library Mus musculus genomic
DEFINITION clone UGCG2M0011N06 F, genomic survey sequence.
ACCESSION A2776986
VERSION A2776986
KEYWORDS GI:12905133
SOURCE GSS.
ORGANISM Mus musculus (house mouse)
REFERENCE Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 455)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
Islam, H., Longacre, S., Mahmood, M., Meenen, E., Pedersen, T.,
Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von
Niederhausern, A., and Wright, D., Weis, R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
CONTACT: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel.: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0011 row: N column: 06
Seq primer: CGTGTAAAGACGACGCCACAT
Class: plasmid ends
High quality sequence stop: 455.
Location/Qualifiers
1..455
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UGCG2M0011N06"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/clone_id="Mouse 10kb plasmid UGCG1M library"
/note="Vector: PMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PMD42 (gi|473214|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to

adapored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

ORIGIN
Query Match 16.2%; Score 81.4; DB 8; Length 455;
Best Local Similarity 68.7%; Pred. No. 2.5e-11;
Matches 112; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

271 TGGGATATCTGCAAGTGCATGAGTGGAGAGGACAGCAATGTGAGGCAAGAGAA 330
45 TGGGGGTCTGTGAACACCACTTCTGACAGATATTAATCTGACAGAGTAAAGAGCA 104
331 TTCTGGCTCAACACAGCTTAGCTCCCTGGTGTGTTCAACTTTGAGAGTTGACACA 390
105 ATCTGCTTATCAACAATGATCTTCTGTTGGTTGTTCAACTTCTATAGTCTGACCCCA 164
391 AGCACTTATTTTGAATATTTAAACAGAGCAACTTTGGG 433
165 GGCATTCAATTTCTGACATATTTTACAGCGAGTGCAATCTGTG 207

RESULT 14
CRL65051 824 bp DNA linear GSS 06-JUN-2004
LOCUS CRL65051
DEFINITION Forward strand read from insert in 3'HPRT insertion targeting and
chromosome engineering clone MHPJ34b03, genomic survey sequence.
ACCESSION CRL65051
VERSION CRL65051.1 GI:49943900
KEYWORDS GSS; genome survey sequence; MICR.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 824)
Adams, D.J., Biggs, P.J., Cox, A.V., Davies, R.M., van der Weyden, L.,
Jonkers, J., Smith, J., Plumb, R.W., Taylor, R.G., Nishijima, I., Yu, Y.,
Rogers, J. and Bradley, A.
Direct Submission
Submitted (20-FEB-2004) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. http://www.sanger.ac.uk/MICR
Location/Qualifiers
1..824
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="MHPJ34b03"
/clone_1lb="MHP"

ORIGIN
Query Match 16.2%; Score 81; DB 9; Length 824;
Best Local Similarity 70.3%; Pred. No. 3.8e-11;
Matches 137; Conservative 0; Mismatches 55; Indels 3; Gaps 2;

270 GTGGGATATCTGCAAGTGCATGAGTGGAGAGGACAGCAATGTGAGGCAAGAGAA 329
86 GCGGAGGTCTGTGAACATCAAGAGCCAGAGATTAATTCATGAGAGAGAAAGAGCA 145
330 ATTCTGCTCAACACAGCTTAGCTCCCTGGTGTGTTCAACTTTGAGAGTTGACACC 389
146 AATCTGCTCAACATGATCTTAGTGAAGCAATGTTGAATCAAACTTGAAGTGAACCA 205
390 AAGCACTTATTTT-GACATATTTAAACAGAGCAACTTTGGAAAAGTTTCTTAT 448
206 GGGGCTTATTTTATGACATATTTAAGACATCAACACTTTGG--AGAAATTTCTGT 263
449 GAAATTAATCAAT 463
264 GACATCAACTCAAT 278

RESULT 15
AK098860/c

LOCUS AK089860 1506 bp mRNA linear HTC 03-APR-2004
DEFINITION Mus musculus activated spleen cDNA, RIKEN full-length enriched library, clone: F830034A08 product: unclassified, full insert sequence.
ACCESSION AK089860
VERSION AK089860.1 GI:26105606
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
JOURNAL MEDLINE 99279253
PUBMED 10349636
REFERENCE
2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komoto, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
JOURNAL MEDLINE 20493374
PUBMED 11042159
REFERENCE
3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Saeki, N., Carninci, P., Komoto, H., Akiyama, J., Nishi, K., Katsunari, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Ogawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system -384-format sequencing pipeline with 384 multichannel sequencer
Genome Res. 10 (11), 1757-1771 (2000)
JOURNAL MEDLINE 20530913
PUBMED 11076861
REFERENCE
4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
JOURNAL MEDLINE 11076861
PUBMED 11076861
REFERENCE
5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
JOURNAL MEDLINE 11076861
PUBMED 11076861
REFERENCE
6 (bases 1 to 1506)
Nature 420, 563-573 (2002)
JOURNAL MEDLINE 11076861
PUBMED 11076861
REFERENCE
7 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Komoto, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, K., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Saeki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-ree@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
COMMENT
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN.

Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC Building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.
Please visit our web site for further details.
URL: http://genome.gsc.riken.jp/
URL: http://fantom.gsc.riken.jp/.

FEATURES

source

1..1506
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strain="NOD"
/db_xref="FANTOM DB:F830034A08"
/db_xref="taxon:10090"
/clone="F830034A08"
/issue_type="activated spleen"
/clone_id="RIKEN full-length enriched mouse cDNA library"
1..1506
misc_feature
/note="unclassified"

ORIGIN

Query Match 16.0%; Score 80; DB 3; Length 1506;
Best Local Similarity 75.0%; Pred. No. 8.5e-11;
Matches 126; Conservative 0; Mismatches 40; Indels 2; Gaps 2;

QY 334 TGGCTCAACACAGCTGCTCCGTGGTGTGTTCAACTTTGAGAGTTGACCAACAGC 393
DB 805 TGGAAACAAACAACTTAATTACTTCTG-ACGTGCAACTCTGAAATCTACCTATAC 747
QY 394 ACTTATTTTGGACATATTTTAAACAGACACAACTTTGGAAAAAGTTCTTTATGAAA 453
DB 746 AGTGA-TTTGACATATATATAACACAGTGCACAACTTGGGAAAAAGTTCTTGTGATGA 688
QY 454 TTATCAATTAAGCTTAAGCATGATCATCTTAAAGCCTTTGCAA 501
DB 687 TTATTACAAACAAAGCTCAAGGCATTAATGAAAGTTCAATTCCAA 640

Search completed: February 10, 2005, 17:33:02
Job time : 2176.78 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 21:17:58 ; Search time 2413.78 Seconds
(without alignments)
10057.285 Million cell updates/sec

Title: US-09-926-375b-7_COPY_11500_12000

Perfect score: 501
Sequence: 1 aaccacacagacatctctc.....gactcatcgcggcttcgttc 501

Scoring table: IDENTITY NUC
Gapop 10_0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

GenEmb1: *
1: gb_Da: *
2: gb_hlg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pac: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_gy: *
13: gb_un: *
14: gb_vl: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	501	100.0	13962	10	MMPSPG X68699 M. musculus
2	501	100.0	17732	6	AX042378 AX042378 Sequence
3	501	100.0	20623	6	AX042372 AL732466 Mouse DNA
4	501	100.0	172535	10	AL732466 Mouse DNA
5	501	100.0	214345	2	AC107726 Mus muscu
6	433.8	86.6	1219	10	MMPSPG X04713 Mouse parot
7	344.2	68.7	236802	2	AC103277 Rattus no
8	299.8	59.8	165313	10	AC125117 Mus muscu
9	284	56.7	284	10	MUSPSPAL M26805 Mouse parot
10	278.4	55.6	236802	2	AC103277 Rattus no
11	238.4	45.6	1733	10	AF153354 Rattus no
12	175.6	35.0	165313	10	AC125117 Mus muscu
13	175.6	35.0	211446	10	AC147845 Mus muscu
14	165.2	33.0	1196	10	AF153355 Rattus no
15	52.4	10.5	161593	9	HSBA49G10 Human DNA
16	45	9.0	62861	2	AC137726 Homo sapi
17	40.2	8.0	134341	9	AC099794 Homo sapi
18	39	7.8	190412	2	CR396584 Danto rer
19	38.8	7.7	21078	4	AT197556 Sus scrofa

20	38.6	7.7	244170	2	AC130152 Rattus no
21	37.8	7.5	160039	9	AC094109 Homo sapi
22	37.8	7.5	160039	9	AC068323 Homo sapi
23	37.8	7.5	195748	10	AL671889 Mouse DNA
24	37.6	7.5	240340	2	AC095354 Rattus no
25	37.6	7.5	270468	2	AC094628 Rattus no
26	37.4	7.5	5519	3	DMMDG3 X95908 D. melanogaster
27	37.4	7.5	74534	3	AC004735 Drosophila
28	37.4	7.5	82912	3	AC004574 Drosophila
29	37.4	7.5	94741	3	AE003607 Drosophila
30	37.4	7.5	159479	3	AC009735 Drosophila
31	37.4	7.5	164712	3	AC007589 Drosophila
32	37.4	7.5	166935	3	AC010920 Drosophila
33	37.4	7.5	169913	3	AC105898 Drosophila
34	37.4	7.5	170675	3	AC008139 Drosophila
35	37.4	7.5	170914	3	AC010843 Drosophila
36	37.4	7.5	179283	3	AC009383 Drosophila
37	37.4	7.5	179396	3	AC022351 Drosophila
38	37.4	7.5	184272	3	AC009211 Drosophila
39	37.4	7.5	185918	3	AC009839 Drosophila
40	37.4	7.5	190801	3	AC009212 Drosophila
41	37.4	7.5	191907	3	AC093157 Drosophila
42	37.4	7.5	203924	2	AC012159 Drosophila
43	37.4	7.5	223505	3	AE003686 Drosophila
44	37.4	7.5	247955	3	AE003698 Drosophila
45	37.4	7.5	296756	3	AE003492 Drosophila

ALIGNMENTS

RESULT 1	MMPSPG	LOCUS	MMPSPG	13962 bp	DNA	linear	ROD 29-OCT-1999
DEFINITION	M. musculus Psp gene for parotid secretory protein.						
ACCESSION	X68699						
VERSION	X68699.1 GI:53809						
KEYWORDS	parotid secretory protein, Mus musculus (house mouse)						
SOURCE	Mus musculus						
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.						
REFERENCE	1 (bases 1 to 13962)						
AUTHORS	Stendesen, P., Laureen, J., Krogh-Pedersen, H. and Hjorth, J.P.						
TITLE	Novel salivary gland specific binding elements located in the PSP proximal enhancer core						
JOURNAL	Nucleic Acids Res. 26 (11), 2761-2770 (1998)						
MEDLINE	98256451						
PUBMED	9592166						
REFERENCE	2 (bases 1 to 13962)						
AUTHORS	Mikkelsen, T.R.						
TITLE	Direct Submision						
JOURNAL	Submitted (07-OCT-1992) T.R. Mikkelsen, Department of Molecular Biology, University of Aarhus, CF Mollers Alle 130, 8000 Aarhus, DENMARK						
FEATURES	Location/Qualifiers						
source	1..13962						
	/organism="Mus musculus"						
	/mol_type="genomic DNA"						
	/strain="C3H/As"						
	/db_xref="taxon:10090"						
	/chromosome="2"						
	/map="Estimate: 69 cm from centromere"						
	/clone="Lambda YP1, Lambda YP3, Lambda YP7"						
	/clone_id="Lambda-PHAGE (Lambda L47.1)"						
	/germline						
	/note="Allele: b"						
enhancer	1..1578						
	/evidence="experimental"						
gene	4632..12433						
	/gene="Psp"						
exon	4632..4678						
	/gene="Psp"						


```

QY      241 GGACCACAATATAGTCACTTGAATCCAGAGATGCTGCAGAGAAACAGATCTCT 300
DB      30863 GGACCACAATATAGTCACTTGAATCCAGAGATGCTGCAGAGAAACAGATCTCT 30922
QY      301 CTCGGTGATCTCTTCAAGCAGTGTAGTCACTCACTAACTGAGTTCATCT 360
DB      30923 CTCGGTGATCTCTTCAAGCAGTGTAGTGTGATCTCACTAACTGAGTTCATCT 30982
QY      361 GAGAGTCCACAGAGAGCTTTGGGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
DB      30983 GAGAGTCCACAGAGAGCTTTGGGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 31042
QY      421 AGTTTGACTTCAACAATTGAGACATAGATGATCTGTCTGACCTAACAACACT 480
DB      31043 AGTTTGACTTCAACAATTGAGACATAGATGATCTGTCTGACCTAACAACACT 31102
QY      481 AGAGCTCATGGGGTTCTGTTT 501
DB      31103 AGAGCTCATGGGGTTCTGTTT 31123

RESULT 5
AC107726 214345 bp DNA linear HTG 03-MAR-2003
LOCUS Mus musculus clone RP24-85E12, WORKING DRAFT SEQUENCE, 8 unordered
DEFINITION pieces.
AC107726 GI:28631318
VERSION HTG; HTGS PHASE1; HTGS DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE
AUTHORS Birren, B., Linton, L., Nussbaum, C., Lander, E., Ali, A., Allen, N.,
TITLE 1 (bases 1 to 214345)
JOURNAL Mus musculus, clone RP24-85E12
AUTHORS 2 (bases 1 to 214345)
REFERENCE
AUTHORS Birren, B., Linton, L., Nussbaum, C., Lander, E., Ali, A., Allen, N.,
TITLE 2 (bases 1 to 214345)
JOURNAL Mus musculus, clone RP24-85E12
AUTHORS 2 (bases 1 to 214345)

```

```

Lindblad-Toh, K., Lu, G., Lu, A., Mabbitt, R., Maclean, C.,
Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,
McDonald, P., Meneu, L., Mihova, T., Mieng, V., Murphy, T., Naylor, J.,
Nguyen, C., Nicol, R., Nord, C., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., Peterson, K., Phunhahang, P., Pierre, N.,
Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C.,
Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,
Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,
Vasilev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X.,
Wyman, D., Young, G., Zainoun, V., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (03-MAR-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 3, 2003 this sequence version replaced gi:28173217.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www.eeg.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 120103
Center clone name: 85 E 12
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 212553 bases at least Q40
Consensus quality: 213324 bases at least Q30
Consensus quality: 213511 bases at least Q20
Insert size: 194000; agarose-fp
Insert size: 213645; sum-of-contigs
Quality coverage: 13.0 in Q20 bases; agarose-fp
Quality coverage: 11.8 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 64960: contig of 64960 bp in length
64961 65060: gap of 100 bp
65061 68275: contig of 3215 bp in length
68276 68375: gap of 100 bp
68376 74568: contig of 6193 bp in length
74569 74668: gap of 100 bp
74669 81021: contig of 6353 bp in length
81022 81121: gap of 100 bp
81122 91450: contig of 10329 bp in length
91451 91550: gap of 100 bp
91551 118992: contig of 27442 bp in length
118993 119092: gap of 100 bp
119093 165215: contig of 46123 bp in length
165216 214345: gap of 100 bp
165217 214345: contig of 49030 bp in length.
Location/Qualifiers
1. 214345
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP24-85E12"
/clone_1ib="RPCT-24 Male Mouse BAC"
1. 64960
/feature="assembly_fragment"
clone_end:SP6
vector_side:left"
misc_feature
65061..68275

```

```
misc_feature /note="assembly_fragment" 68376. 74568
misc_feature /note="assembly_fragment" 74659. 81021
misc_feature /note="assembly_fragment" 81122. 91450
misc_feature /note="assembly_fragment" 91551. 118992
misc_feature /note="assembly_fragment" 119093. 165215
misc_feature /note="assembly_fragment" 165316. 214345
misc_feature /note="assembly_fragment clone_end:r7
vector_side:right"
```

ORIGIN

Query Match	100.0%;	Score 501;	DB 2;	Length 214345;
Best Local Similarity	100.0%;	Pred NO 4	2e-145.	

Oy	1	AACCCACAGAAACCTTTCCATTTGATTCATTTGATTCCTTTCTTAACAAATCTGAA	60
Db	61415	AACCCACAGAAACCTTTCCATTTGATTCATTTGATTCCTTTCTTAACAAATCTGAA	61474
Oy	61	CTGACCCCTTGTGGAGCTCAAGTTTGA CAACATAGGGCTTTGAACTTGGCAAGGT	120
Db	61475	CTGACCCCTTGTGGAGCTCAAGTTTGA CAACATAGGGCTTTGAACTTGGCAAGGT	61534
Oy	121	CCATCACTGTCAACCCAGCATCTGGGTGACCTTTGGGTGGAATATCTTGGCTAACCTT	180
Db	61535	CCATCACTGTCAACCCAGCATCTGGGTGACCTTTGGGTGGAATATCTTGGCTAACCTT	61594
Oy	181	AGATATTTCTTGGAGATCTTTAGAACATCCAGGAAATAGGGCTTGATTCATCCTG	240
Db	61595	AGATATTTCTTGGAGATCTTTAGAACATCCAGGAAATAGGGCTTGATTCATCCTG	61654
Oy	241	GGACCAACAATTAAGTCAACCTAGATCCAGAGAGATCTGTGACAGAGAAACAAGATCTCT	300
Db	61655	GGACCAACAATTAAGTCAACCTAGATCCAGAGAGATCTGTGACAGAGAAACAAGATCTCT	61714
Oy	301	CTCGTGTGCATCTTCTTCAAAAGCAGTAGTAGTGA CTCACCTAAA CTGAGTTCCCATCT	360
Db	61715	CTCGTGTGCATCTTCTTCAAAAGCAGTAGTAGTGA CTCACCTAAA CTGAGTTCCCATCT	61774
Oy	361	GAGAGTCCACAGAGAGCTTTGGGGCAAGAAAGCAGAGGGAAGGCACTGTTTGTGTGTAA	420
Db	61775	GAGAGTCCACAGAGAGCTTTGGGGCAAGAAAGCAGAGGGAAGGCACTGTTTGTGTGTAA	61834
Oy	421	AGTTTGACTTAACAATTTGGAACATAGATGACATTTGTGCAGACTTAACAACAACCT	480
Db	61835	AGTTTGACTTAACAATTTGGAACATAGATGACATTTGTGCAGACTTAACAACAACCT	61894
Oy	481	AGACTCAATGGGTTCGTTT	501
Db	61895	AGACTCAATGGGTTCGTTT	61915

LOCUS	1219 bp	DNA	linear	ROD 07-JUN-1987
DEFINITION	Mouse parotid secretory protein (PSP) gene exon 1 and flanks.			
ACCESSION	X04713			
VERSION	X04713.1	GI:53808		
KEYWORDS	parotid secretory protein.			
SOURCE	Mus musculus (house mouse)			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 1219)			
AUTHORS	Shaw, P. and Schibler, U.			
TITLE	Structure and expression of the parotid secretory protein gene of mouse			

JOURNAL	J. MOL. BIOL.	192 (3),	567-576 (1986)
MEDLINE	87169741		
PUBMED	3560226		
COMMENT	Determination of PSP gene structure revealed eight introns, ranging from 20 to 200 bp. The first exon (270b)		

FEATURES	location/Qualifiers
source	1..1219
	/organism="Mus musculus"
	/mol_type="genomic DNA"
	/db_xref="taxon:10090"
	/tissue_type="parotid gland"
repeat_unit	131..136
	/note="inverted repeat A"
repeat_unit	137..142
	/note="inverted repeat A"
misc_feature	253..260
	/note="polyoma enhancer-like sequence"
misc_feature	473..480
	/note="SV40 enhancer-like sequence"
misc_feature	587..594
	/note="hpc. DNA bending sequence"
promoter	821..828
	/note="CAAT motif"
promoter	875..881
	/note="TATA motif"
mRNA	906..947
	/note="exon 1 (noncoding)"
intron	948..>1219
	/note="intron I"

Query Match	86.6%	Score 43.8;	DB 10;	Length 1219;
Best Local Similarity	95.4%;	Pred. No. 5.2e-124;		
Matches 480; Conservative	0;	Mismatches 17;	Indels 6;	Gaps 3

Oy	I	AAOCCACAGAACATCTTCCATTTGATATTCATTTCCGATCTCTTCTAGACAAATACGAA	60
Db	627	AAOCCACAGAACATCTTCCATTTGATATTCGGAATCTTCTTAGACAAATCTGAA	686
Oy	61	CTGACCCCTTGTTGGCAGCTCTCAAGTTTGAACAACATATGGGCTTTGAACTTGGCCAAAGT	120
Db	687	CTGACCCCTTGTTGGCAGCTCTCAAGTTTGAACAACAATAGGCTTTGCAACTTGGCAACAAGT	746
Oy	121	CCATCACTGTCAOCCCAAGCATCTGGGTGACCTTTGGTTGGAATATCTTGGCTPAACCTT	180
Db	747	CCATCACTGTCAOCCCAAGCATCTGGGTGACCTTTGGTTGGAATATCTTGGCTPAACCTT	806
Oy	181	AGATA-TTTCCTTTGGAGTATCTTTAGAACATCCAGAAATAGGCTTGTATCTTCATCTT	239
Db	807	AGATATTTTCTTGGAGTATCTTTAGAACATCCAGAAATAGGCTTGTATCTTCATCTT	866
Oy	240	GGAGCCACAAATATATAGTCAOCCCTAGATCCOAGAGATCGT-GCAGAGAAACAAGATCT	298
Db	867	GGAGCCACAAATATATAGTCAOCCCTAGATCCOAGAGATCGTGCAGAGAAACAAGA---	923
Oy	299	CTCTCAGTGCATCTCTTCTCAAGCAGTAGTAGTCACTTCACTPAACTGAGTTCCCAT	358
Db	924	-TCTCGTGTGATCTCTTCTCAAGCAGTAGTAGTCACTTCACTPAACTGAGTTCCCAT	982
Oy	359	CTGAGAGTCCAAGAGAGCTTTGGGGCCAGAGCAGAGGAGGACACTGTTTGTGTGCTGT	418
Db	983	CTGAGAGTCCAAGAGAGCTTTGGGGCCAGAGCAGAGGAGGACACTGTTTGTGTGCTGGG	1042
Oy	419	AAAGTTTGACTCTPAACAAAATTTGAACAACATAGATGCACTTGTGTGAGCTTAAACAAC	478
Db	1043	AAAGTTTGACTCTPAACAAAATTTGAACAACATAGATGATTTGTGTGAGCTTAAACAAC	1102
Oy	479	CTAGACTCAGTGGGTTCTGTTT 501	
Db	1103	CTAGACTCAGTGGGTTCTGTTT 1125	

```

RESULT 7
AC103277/c
LOCUS
DEFINITION
Rattus norvegicus clone CH230-23561, *** SEQUENCING IN PROGRESS
*** 2 unordered pieces.
AC103277
236802 bp DNA linear HTG 13-MAY-2003
AC103277.6 GI:30579004
HTG; HTGS PHASE1; HTGS DRAFT; HTGS ENRICHED.
KEYWORDS
Rattus norvegicus
SOURCE
Rattus norvegicus
Rattus norvegicus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 236802)
Muzny,D,Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alpbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Bismail,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Bunay,C., Burch,P., Butrell,K., Caesar,H., Cederon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Deigado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Georgievski,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunatirup,P., Haaland,W., Hamill,C., Hamilton,K., Hernandez,J.,
Harvey,T., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogue,M.,
Hollins,B., Howells,S., Huylk,S., Hume,J., Idlebird,D., Jackson,A.,
Jacks,J., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolyvet,A.,
Karpeth,S., Kelly,S., Kelly,S., Khan,Z., King,Z., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorenshuwa,L., Loulseg,H., Lozada,R.J., Lu,X., Ma,U.,
Maheshwari,M., Mahindartine,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapa,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Mijic,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervils,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K.,
Pawolemech,O., Okunomu,G., Olapumunsegon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfankoch,C.,
Piopier,F., Polindexter,A., Popovic,D., Primus,E., Pu,L.,-L.,
Pizarro,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richard,S., Riggs,F.,
Rivers,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Saverly,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Sma's,D.,
Speed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Soosa,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villaseña,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausen,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
TITLE
Direct Submission
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 236802)
Worley,K.C.
TITLE
Direct Submission
JOURNAL
Submitted (24-NOV-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
REFERENCE
3 (bases 1 to 236802)

```

```

AUTHORS
Rat Genome Sequencing Consortium.
TITLE
Direct Submission
JOURNAL
Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
COMMENT
On May 13, 2003 this sequence version replaced gi:25086871.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both ends and sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: G4RO
Center clone name: CH230-23561
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 226151 bases at least Q40
Consensus quality: 228208 bases at least Q30
Consensus quality: 228980 bases at least Q20
Estimated insert size: 234093; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 235405: contig of 235405 bp in length
* 235406 235505: gap of unknown length
* 235506 236802: contig of 1297 bp in length.
FEATURES
source
1..236802
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-23561"
misc_feature
1..3075
/feature="wgs_end_extension
clone_end:Sp6"
misc_feature
complement(5172..5970)
/feature="clone boundary
clone_end:Sp6"
misc_feature
site:BCORI
end_sequence:82095899"
228453..229719
/feature="wgs_contig"
233558..233880
/feature="clone_boundary
clone_end:T7
site:BCORI
end_sequence:82095897"
ORIGIN
Query Match 68.7% Score 344.2; DB 2; Length 236802;
Best Local Similarity 86.8% Pred No 3.9e-96;
Matches 440; Conservative 0; Mismatches 53; Indels 14; Gaps 5;

```


repeat_region	56922..57044	/rpt_family="B4"
repeat_region	57783..57888	/rpt_family="B2"
repeat_region	58600..58763	/rpt_family="L1"
repeat_region	58764..58853	/rpt_family="RRVK"
repeat_region	64713..64893	/rpt_family="L1"
repeat_region	64951..65149	/rpt_family="L1"
repeat_region	65128..65337	/rpt_family="L1"
repeat_region	66707..66851	/rpt_family="RRV1"
repeat_region	66855..66923	/rpt_family="MALR"
repeat_region	67272..67468	/rpt_family="B2"
repeat_region	68460..68569	
Query Match	59.8%;	Score 299.8; DB 10; Length 16533;
Best Local Similarity	83.5%;	Pred. No. 2.9e-82;
Mismatches 420; Conservative	0;	Mismatches 62; Indels 23; Gaps 7;

	Query Match	59.8%	Score 299.8 ; DB 10	Length 165113;
	Best Local Similarity	83.5% ;	Pred. No. 2.9e-87 ;	
	Matches 430; Conservative	0 ; Mismatches 62 ; Indels 23 ; Gaps 7 ;		
b		3	CCCAACAGAACTTCATCATGTGATTCTTCTTGAGACTCATAAACAATCGTAAGTT	62
y		2500	CCCCAACAAATTCCTCATTCGCATTCGGCTTCCTTAGCACAGAACAATCTGAATT	24941
b		63	GACCCTTGTTGGCAGTCTCAAAGTTTGACAAACAATAGGCCCTTTG-----AACTTGGCA	114
y		24940	GA-CATATTTTGGAGCCTCAGGTTTTCCAATCAGAAAATCAGTCYTTGSCAGAAATTCCT	24887
b		115	CAGAGTCATCACTGTCACCCAGACATCCTGGGTGACCTTTGGGTTGGAATATCTTGAGCT	174
y		24881	CTCATCCACCACTGAATCTTAAGTATTCAGTGAACCTTTGGTTGGAATATCTTGAGCT	24822
b		175	AACCTGATATATTTCTTTGGAGTATCTTTGAACATCCAGAAATAGSGCTGATTCTC	234
y		24821	AACCTTAGATATTTCTTTGG-GTCTTTTGAACATCCGGAATATGGATCTGATTCTC	24763
b		235	ATCTTGGGACCAATAT-- -- -AGTACCCTTAATCCCGAGGATCGTGCAGAG	287
y		24762	ATCTCGGGACCAATATPAAGGCCACAGTACCCCAGAAATCCAGGAGATCGTGCAGAG	24700
b		288	AAACAAGATCTCTCGTGTGATCCTCTTCAAAGCAGTGAATGATCTCCACTAAAC	347
y		24702	AACAAGAAAA---TCTCGTGTGATCTCTTCAAAGCAGTGAATGATCTCCACTAAAC	24644
b		348	TGAGTCCCATCTGAGATCCACAGAGGCTTTGGGCAAGAACAGAGGAAGGACCTG	407
y		24646	TGAGTCCCATCT-TTAAAGCTTCAAGAGGCTTTGGGCAAGAAATGAGAGGAAGGCGCTG	24588
b		408	TTTGTGTGTGAAGTTTGAATCT-CTAACAAATTTGAAGACATGAGATGATGTGTCAAG	466
y		24587	TTTGTGTGTGAAGACGTTTGAATGTTAACAAATTTGAAGACATGAGATGATGTGTCAAG	24520
b		467	ACTAACACACCTAGACTCATGTGGGTCTGTTT	501
y		24527	ACTAACACACCAAGACTCATGTGGGTCTGTTT	24493
MUSPSPAL				
MUSPSPAL				
LOCUS				
DEFINITION			284 bp DNA linear ROD 20-MAY-1994	
ACCESSION			Mouse parotid secretory protein (PSP) gene, exon 1, clone YPIB2.8.	
VERSION			M26805	
KEYWORDS			M26805.1 GI:200554	
SOURCE			parotid secretory protein; secretory protein.	
ORGANISM			1 of 3	
			Mus musculus (house mouse)	
			Mus musculus	

121 CCATCACTGTCACCCAGCATCCTGGGTGACCTTGGGTGGAATATCTTGGCTAACCTT

Db 1527 CCATCACTGTACTCAAGCATCTCGGTGACCTTTGGTTGGATATCTTGGCTTAACCTT 1586
 |||||
 Qy 181 AGATATTTTCTTTGGAGATCTTTAGAACATCCAGAAATAGGGCTTGATTTCTATCTCG 240
 |||||
 Db 1587 AGGATATTTCTTTGGAGTCTCTTTA-TTCATTTCCGGAATAGGGCTTGTTCTATCTTG 1645
 |||||
 Qy 241 GGACCACATATTA-----GTCACTCTAGATCCAGAGATCGTCAGAGAAACAG 293
 |||||
 Db 1646 GGACCACGATTAAGGCCAGGTCACTCCCAAAATCCAGAGATCGTCAGAGAAACAG 1705
 |||||
 Qy 294 GATCTCTCTCGTGCATCTTTCTTCAAGA 325
 |||||
 Db 1706 G-----CTTCTGTGCTCTCTTTCTCCAAACGA 1733
 |||||


```

repeat_region      /rpt_family="L1"
                    28037..28164
repeat_region      /rpt_family="Alu"
                    30431..31273
repeat_region      /rpt_family="ERVK"
                    31535..31609
repeat_region      /rpt_family="B4"
                    31833..31922
repeat_region      /rpt_family="Alu"
                    31962..32291
repeat_region      /rpt_family="ERVK"
                    32338..33367
repeat_region      /rpt_family="ERVK"
                    37527..37856
repeat_region      /rpt_family="ERVK"
                    37975..38424
repeat_region      /rpt_family="RMR12"
                    38425..38774
repeat_region      /rpt_family="L1"
                    39296..39555
repeat_region      /rpt_family="RMR12"
                    40214..42773
repeat_region      /rpt_family="L1"
                    42977..43025
repeat_region      /rpt_family="L1"
                    43015..43516
repeat_region      /rpt_family="L1"
                    45272..45464
repeat_region      /rpt_family="B4"
                    51509..51627
repeat_region      /rpt_family="ERV1"
                    51836..52252
repeat_region      /rpt_family="L1"
                    52249..52290
repeat_region      /rpt_family="L1"
                    52806..52922
unsure             /note="Sequence derived from one plasmid subclone."
repeat_region      52864..52951
repeat_region      /rpt_family="L1"
                    52953..53119
repeat_region      /rpt_family="B2"
                    53160..53479
repeat_region      /rpt_family="L1"
                    53530..53636
repeat_region      /rpt_family="L1"
                    53637..54191
repeat_region      /rpt_family="ERVK"
                    54515..54872
repeat_region      /rpt_family="L1"
                    55568..55649
repeat_region      /rpt_family="B4"
                    56922..57044
repeat_region      /rpt_family="B4"
                    57783..57888
repeat_region      /rpt_family="B2"
                    58600..58763
repeat_region      /rpt_family="L1"
                    58764..58853
repeat_region      /rpt_family="ERVK"
                    64713..64893
repeat_region      /rpt_family="L1"
                    64951..65149
repeat_region      /rpt_family="L1"
                    65128..65337
repeat_region      /rpt_family="L1"
                    66707..66851
repeat_region      /rpt_family="ERV1"
                    66855..66923
repeat_region      /rpt_family="MALR"
                    67272..67468
repeat_region      /rpt_family="B2"
                    68460..68569

```

```

Query Match      35.0%; Score 175.6; DB 10; Length 165313;
Best Local Similarity 73.2%; Pred. No. 1.8e-43;
Matches 284; Conservative 0; Mismatches 89; Indels 15; Gaps 4;

QY      122 CATCACTGTGACCCAGACATCTCTGGGTGACCTTTGGTTGGAATATCTTGCTTAACCTTA 181
      |||||
Db      46859 CATCAAAACCATTCATATGCTTGGGTGACCTTTGGTTGGAATATGTTGCT-ACCTTA 46917

QY      182 GATATTTCTTTGAGATCTCTTTGAACATCCAGAAATAGGGCTTGAATCTTCATCTCG 241
      |||||
Db      46918 GGTATTACTTTGAGCTCTTTTGAACATCTCTTGAACAGACTTGTCTTCATCTCGT 46977

QY      242 GACCACATAT-----AAGTCAACCTTAGAATCCAGAGATCGTGCAGAGAAACA 292
      |||||
Db      46978 GACCTCATATATATAAAGCCAGATCCTTAGATTCAGAGTCCAGAGTCCGTGTAATAAACA 47037

QY      293 GGATCTCTCTGTGTGATCTCTTTCAAAGACATGATGACTCCACTTAACTGAGT 352
      |||||
Db      47038 GGA-----TTTCAATGATGATCAATTTTCAACCTCAATAGGGATCCCTGACTGTGA-C 47092

QY      353 TCCCATGTGAGATGCAGAGAGGCTTTGGGCAAGACAGAGAGGAGCACTGTGT 412
      |||||
Db      47093 ACACATCTGAGGGCTTTTGAAGAGCTTTAGAACAGAGACAGGAGAGCTTGTGT 47152

QY      413 GTTGTAAAGTTTGTACTCTTAACAATTTGAAGACATGATGACTGTGTCAACTAAC 472
      |||||
Db      47153 GTTGGGATGATATTATTAACAATAATTTGAAGGATGATTAATCTGTGCAGACTAAT 47212

QY      473 AACAACTTAGACTCATGTGGGTCTGT 500
      |||||
Db      47213 GACAATCAAGTTTCAATGTGTCTGT 47240

RESULT 13
AC147545/C      211446 bp      DNA      linear      ROD 08-SEP-2004
LOCUS      Mus musculus chromosome 14 clone RP23-355L20, complete sequence.
DEFINITION      AC147545
ACCESSION      AC147545.3 GI:51921462
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 211446)
REFERENCE
AUTHORS      Wilson,R.K.
TITLES      The sequence of Mus musculus clone
JOURNAL      Unpublished
2 (bases 1 to 211446)
REFERENCE
AUTHORS      Wilson,R.K.
TITLES      Direct Submision
JOURNAL      Submitted (10-DEC-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 211446)
REFERENCE
AUTHORS      Wilson,R.K.
TITLES      Direct Submision
JOURNAL      Submitted (28-MAY-2004) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
4 (bases 1 to 211446)
REFERENCE
AUTHORS      Wilson,R.K.
TITLES      Direct Submision
JOURNAL      Submitted (08-SEP-2004) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
COMMENT
On Sep 8, 2004 this sequence version replaced gi:4777607.

```

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: W_BA0355L20

```

FEATURES
source Location/Qualifiers
1..211446
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="14"
/clone="RP23-355L20"

ORIGIN
Query Match 35.0%; Score 175.6; DB 10; Length 211446;
Best Local Similarity 73.2%; Pred. No. 1.7e-43;
Matches 284; Conservative 0; Mismatches 89; Indels 15; Gaps 4;

QY 122 CATCACTGTCAACCAAGCATCTCGGAGCTTTGGGTTGGATATCTGGCTTAACCTTA 181
197638 CATCAAAACCATTCATAGCTTGGGAGACCTTGGTTGGATATGTTGGCT-ACCTTA 197580
DB 182 GATATTTCTTTGAGATCTTGAACATCCAGGAATAGGGCTGATCTCATCTCG 241
197579 GGTATTTACTTGAGCTCTTTTGAACATCTGTAACCAAGCTTGGCTTCTCATCTCT 197520
QY 242 GACCAACAATAT-----AGTCACTTGAATCCAGAGATCGTGCAGAGAA 292
197519 GACCTCAATATATAAGCAAGTCACTTGAATCCAGAGATCGTGCAGAGAA 197460
QY 293 GATCTCTCTGTCGATCTCTTCAAGCAAGTGAATGATCTCACTCAATAGT 352
197459 GGA-----TTTCATGTCATATTTTCAAAACCTCAAAATAGGGATCCCTCACTGTA-C 197405
QY 353 TCCATCTGAGTCCACAGAGGCTTTGGGGCAAGAGAGAGGAGGACATGTTGT 412
197404 ACAATCTGAGGGCTTCAAGAGCTTTGAACAAGAGACAGGAGGTTCACTTTGT 197345
QY 413 GTTGTAAGTTTGACTTACCAATTTGAAGACATAGATGACATTTGTCAAGACTAC 472
197344 GTTGGGATATATTTACATACAAATTTGAAGGATGATTAATCTGTCAAGACTAAT 197285
QY 473 AACCACTAGACTCATGTGGTCTGT 500
197284 GACATCAAGGTTTCATGTGTTCTGT 197257
DB

RESULT 14
AF153355 1196 bp DNA linear ROD 29-FEB-2000
LOCUS
DEFINITION Rattus norvegicus Smgb gene, upstream sequence and exon 1.
ACCESSION AF153355
VERSION AF153355.1 GI:5031464
KEYWORDS
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Bukaryaota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 1196)
Gupta,N., Asp,E., Levan,G. and Mirels,L.
Structure and chromosomal localization of the rat salivary Psp and
Smgb genes
Gene 243 (1-2), 11-18 (2000)
JOURNAL MEDLINE 20156164
PUBMED 10675608
REFERENCE 2 (bases 1 to 1196)
AUTHORS Mirels,L.
TITLE Direct Submission
JOURNAL Submitted (21-MAY-1999) Molecular and Cell Biology, University of
California, Berkeley, 401 Barker Hall #3204, Berkeley, CA
94720-3204, USA
FEATURES
source Location/Qualifiers
1..1196
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"

ORIGIN
Query Match 33.0%; Score 165.2; DB 10; Length 1196;
Best Local Similarity 76.9%; Pred. No. 3.9e-40;
Matches 260; Conservative 0; Mismatches 58; Indels 20; Gaps 4;

QY 3 CCCACAGAACATCTTCCATGATTCATCTTGATCTTCTTCTTGACAAATTCGAAT 62
864 CACACACAAATTCCTCCATTCATCTGAGGCTTCTTCCAGAAATAGCAACT 923
DB 63 GACCCCTTGTCAGTCTCAAGTTGACAAATAGGGCTTG-----ACCTGGCA 114
QY 924 GACATCTTGACCTCAGCTCAGTCTCACTCAAGAAATCACTTGGCAGAACTTCT 982
115 CAAGTTCATCACTGTCAACCAAGATCTGAGTGAACCTTGGTTGAATATCTTGCT 174
DB 983 CTCATTCATCATTTGTCATCAAGATCATTTGGTGAACCTTGGATTTGATCTTGCT 1042
QY 175 AACCTTAGATATTTCTTGGATATCTTGAACATCCAGAAATAGGGCTTGAATCTC 224
1043 AACCTTAGATATTTCTTGGATATCTTGAACATCCAGAAATAGGGCTTGAATCTC 1102
DB 235 ATCTTGAGACCAATATATA-----GTCACTTGAATCCAGAGATCTGTCAGAGA 287
QY 1103 ATCTTGAGACCAATATATA-----GTCACTTGAATCCAGAGATCTGTCAGAGA 1162
DB 288 AACAGGATCTCTCTGTCATCTTCTTCAAGCA 325
1163 AACAGG-----CTTCTGTCTCTCTTCCAAACA 1196
DB

RESULT 15
HSBA49G10 161593 bp DNA linear PRI 20-JUN-2001
LOCUS
DEFINITION Human DNA sequence from clone RP11-49G10 on chromosome 20. Contains
a gene similar to bovine salivary protein BSP30, the LOC51297 gene
for LUNX protein, 5' end of a gene encoding a protein similar to
murine von Ebner minor salivary gland protein, a novel gene, a
pseudogene similar to STAT-induced STAT inhibitor-2, a pseudogene
similar to ribosomal protein L12, a putative novel transcript,
ESTs, STSs, GSSs and a CpG island, complete sequence.
ACCESSION AL121901
VERSION AL121901.20 GI:8249854
KEYWORDS HTG; BSP30; CpG island; LOC51297; LUNX; RP12; STAT-inhibitor.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Bukaryaota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 161593)
REFERENCE Tracey,A.
JOURNAL Submitted (14-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jun 5, 2000 this sequence version replaced gi:8017404.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
EM, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEB; Information

on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr20>
IMPORTANT: This sequence is not the entire insert of clone RP11-49G10. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.
The true left end of clone RP11-49G10 is at 1 in this sequence. The true left end of clone RP5-1187J4 is at 161494 in this sequence. The true right end of clone RP4-7330J3 is at 27823 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP11-49G10 is from the library RPCT-11.1 constructed by the group of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
VECTOR: pBACe3.6.

```
FEATURES
source
    1..161593
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
        /chromosome="20"
        /clone="RP11-49G10"
        /clone_1lb="RPCT-11.1"
        /complement(1..107)
        /note="match: GSS: Em:AQ607627"
        6..706
        /note="match: GSS: Em:AQ051854"
        137..548
        /note="match: GSS: Em:AQ550113"
        530..603
        /note="37 copies 2 mer ca 66% conserved"
        533..686
        /note="2 copies 77 mer 83% conserved"
        613..678
        /note="33 copies 2 mer ca 75% conserved"
        3030..3174
        /note="LIPB3 repeat: matches 5941..6084 of consensus"
        5140..5183
        /note="22 copies 2 mer tt 75% conserved"
        5184..5344
        /note="LIPB2 repeat: matches 5986..6146 of consensus"
        /complement(6104..6463)
        /note="match: GSS: Em:AQ339300"
        6120..6239
        /note="2 copies 60 mer 85% conserved"
        /complement(8476..9002)
        /note="match: GSS: Em:AQ537190"
        /complement(14766..15061)
        /note="match: STS: Em:HS922114S"
        14772..14779
        /note="Tandem repeat. Forced join. Approximately 500 bases missing according to restriction digest."
        16219..16387
        /note="match: GSS: Em:AL001671"
        16225..16387
        /note="match: GSS: Em:AL013146"
        18325..18826
        /note="match: GSS: Em:AQ458133"
        18328..18828
        /note="match: GSS: Em:AQ53957"
        18333..18770
        /note="match: GSS: Em:AQ765930"
        /complement(20975..21245)
        /note="match: GSS: Em:A2102442"
        20999..21238
        repeat_region
```

```
misc_feature
    21031..21263
        /note="4 copies 60 mer 69% conserved"
        /note="match: GSS: Em:A2055779 Em:A2084220"
        21144..21225
        /note="41 copies 2 mer ag 75% conserved"
        /complement(21172..21428)
        /note="match: GSS: Em:AQ395327"
        /complement(23473..23950)
        /note="match: GSS: Em:AQ727877"
        /complement(23637..23948)
        /note="match: GSS: Em:AQ691469"
        23950..24335
        /note="match: GSS: Em:AQ152593"
        23955..24456
        /note="match: GSS: Em:AQ508832"
        26979..27422
        /note="match: GSS: Em:AQ636646"
        31882..32061
        /note="3 copies 60 mer 73% conserved"
        31921..32032
        /note="56 copies 2 mer ct 63% conserved"
        33693..33762
        /note="MER47 repeat: matches 2251..2322 of consensus"
        33733..33785
        /note="MER47 repeat: matches 1..55 of consensus"
        35535..48787
        /note="BA49G10.1"
        /join(35535..35585,36507..36678,40308..40452,41455..41562,42783..42935,45523..45604,46980..47043,47896..47936)
        /gene="BA49G10.1"
        /product="BA49G10.1 (similar to bovine salivary protein BSP30)"
        /note="match: CDNAS: Em:U79414 Em:U79413 Em:AX061621"
        /note="match: ESTs: Em:A1654622"
        /evidence="not experimental"
        /join(36522..36678,40308..40452,41455..41562,42783..42935,45523..45604,46980..47043,47896..47936)
        /gene="BA49G10.1"
        /note="match: proteins: Tr:P79125 Sw:P07743"
        /codon_start=1
        /evidence="not experimental"
        /product="BA49G10.1 (similar to bovine salivary protein BSP30)"
        /protein_id="CAC03546.1"
        /db_xref="GI:9801234"
        /db_xref="UniProt/TREMBL:Q9BQ00"
        /translation="MLQMKVLLCGVITGTSSESLDNLGNDISVYNDKLEPVLEHGL
        EIVDNTLKGILEKLVKVDGLVQKSSAWQLAKQAEKELNNVSKILPTTDFGL
        KTSNGLIDVKAEPIDDKGINTLSPPVANVYVAGPIIGQIINLKASLDLPAVETP
        DQTHOPVAVLGECSADPTSLISLIDKHSQINKFVNSVINTKSTVSSILQKEICP
        LIRIRHSDVNVIQVVDNPKHTQLQTLI"
        38155..38751
        /note="LIPB8 repeat: matches 4514..5125 of consensus"
        38750..39785
        /note="LIPB8 repeat: matches 5128..6163 of consensus"
        48769..48774
        /gene="BA49G10.1"
        48787
        /gene="BA49G10.1"
        52540..53450
        /note="CpG island"
        /evidence="not experimental"
        53831..54141
        /note="match: STS: Em:G31731"
        /complement(53831..54131)
        /note="match: GSS: Em:AQ97745"
        53842..53969
        /note="match: GSS: Em:AQ429347"
        /complement(53844..54141)
        /note="match: GSS: Em:AF101963"
        53858..54022
        /note="match: GSS: Em:AQ392891"
        misc_feature
```

```
misc_feature      complement(53908..54141)
                   /note="match: GSS: Em:AQ415287"
misc_feature      53958..54139
                   /note="match: GSS: Em:AQ378081"
misc_feature      53970..54139
                   /note="match: GSS: Em:AQ527291"
misc_feature      53972..54142
                   /note="match: GSS: Em:AQ429347"
misc_feature      54000..54141
                   /note="match: GSS: Em:AQ392510"
misc_feature      complement(54027..54139)
                   /note="match: GSS: Em:AQ533658"
misc_feature      54030..54138
                   /note="match: GSS: Em:AG000935"
misc_feature      54041..54149
                   /note="match: GSS: Em:AQ392891"
gene              54767..55668
                   /gene="bA49G10.2"
                   /pseudo
CDS              54767..55668
                   /gene="bA49G10.2"
                   /note="bA49G10.2 (similar to STAR-induced STAR
inhibitor-2)
```

```
Query Match      10.5%; Score 52.4; DB 9; Length 161593;
Best Local Similarity 60.1%; Pred. No. 5.2e-05;
Matches 197; Conservative 0; Mismatches 91; Indels 40; Gaps 5;
```

```
QY      135 CAAGCATCTGGGTGACCTTTGGGTGGAATATCTTGCTAACCTTA-----G 182
DB      35364 CAAGCATCTGGGTGACCTTTGGGTGGAATATCTCTCTGCTCATCTCGAGAGGGG 35423
QY      183 ATATTTCTTTGGAGTATCTTTAGACATCCAGGAAATAGGCTTGATCT----- 233
DB      35424 TTATTTCTTTGGAGTATCTTTAGATGATCCAGAACCAAGGCTTGCTTGCGGAGC 35483
QY      234 ----CATCTGGGACCAATATATAGTACCC-----TAGAATCCAGAGATCG 279
DB      35484 CTTCATCTCCATCTCCACATATATAGCCGCCACATTCAGCAGAAATCCAGCAGACTG 35543
QY      280 TGCAGAGAAACAAGATCTCTCGTGTGATCTCTTCAAGCAGTGAAGTGAATC 339
DB      35544 TGCATGGGGCAAGGA---TTTATAGCATCTCTCTTAACCGCTGAGTGGCGTCT 35599
QY      340 CACTAACTGAGTTCATCTGAGAGTCCACAGAGGCTTT-GGGCAAGAACGAGAGG 398
DB      35600 CCCCCAGTGGCACCACACGCTGCTTTTGAAGGCTTTGGGGCAAGAGCAGAGGG 35659
QY      399 AAGGACTGTTTGTGTGTAAGTTT 426
DB      35660 GTGGGATTGCTTGGGTAGCTAATATTT 35687
```

```
Search completed: February 10, 2005, 14:09:34
Job time : 2418.78 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 22:02:33 ; Search time 2171.78 Seconds
(without alignments)
8780.927 Million cell updates/sec

Title: US-09-926-375b-7_COPY_11500_12000

Sequence: 501 1 aaccacaacagacatctctc.....gactcatggtggtctctt 501

Scoring table: IDENTITY NUC
Gapop 10_0, Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database:

EST.*
1: gb_est1.*
2: gb_est2.*
3: gb_hic.*
4: gb_est3.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_g881.*
9: gb_g882.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	127	25.3	747	4	BG865470 602783596
2	127	25.3	865	4	BG871618 602790109
3	127	25.3	866	4	BG869398 602788759
4	126	25.1	767	4	BG867661 602787451
5	125	25.0	680	4	BH171830 imagec 5
6	125	25.0	744	4	BG865845 602788261
7	125	25.0	938	4	BG867523 602788460
8	124	24.8	832	4	BG869852 602789378
9	122	24.4	1007	4	BG867502 602788430
10	120	24.0	754	4	BG873122 602794520
11	119	23.8	249	1	AA562244
12	117.6	23.5	871	6	CB575686
13	115	23.0	563	8	AG974026
14	115	23.0	671	2	BF301350 602029803
15	115	23.0	827	2	BF536340 602051990
16	114.4	22.8	442	1	AA798080 vx67a05.r
17	113	22.6	686	4	BG867018 602786008
18	113	22.6	920	4	BG865620 602783777
19	108	21.6	918	2	BF535796 602051107
20	107.6	21.5	880	2	BF300134 602032028
21	107.4	21.4	955	2	BF540271 602052325
22	105.4	21.0	771	2	BF537617 602053392
23	105	21.0	848	4	BG869732 602787559
24	104	20.8	903	2	BF302106 602033055

25	103	20.6	726	2	BF302348	BF302348 602031330
26	99.4	19.8	781	2	BF538311	BF538311 602053793
27	99	19.8	539	1	AA930330	AA930330 v859b09.r
28	99	19.8	844	2	BF300497	BF300497 602031631
29	89	17.8	850	2	BF302185	BF302185 602033141
30	82.8	16.5	873	2	BF535056	BF535056 602054316
31	80.4	16.0	365	9	CR145142	CR145142 Forward s
32	74	14.8	797	1	AA726504	AA726504 v092f01.r
33	57	11.4	813	2	BF536339	BF536339 602051989
34	57	11.4	996	2	BF301924	BF301924 602032952
35	55	11.0	843	4	BG870730	BG870730 602722131
36	49	9.8	784	4	BG871734	BG871734 602791969
37	49	9.8	992	4	BG870291	BG870291 602791160
38	47.6	9.5	428	1	AI930435	AI930435 u164603.y
39	47	9.4	395	6	BY791312	BY791312 BY791312
40	47	9.4	357	6	BY776189	BY776189 BY776189
41	47	9.4	972	3	AK009032	AK009032 Mus muscu
42	47	9.4	972	3	AK010181	AK010181 Mus muscu
43	46	9.2	461	1	AI574478	AI574478 uK21f11.y
44	46	9.2	625	1	AI151916	AI151916 u147f07.y
45	44	8.8	1046	4	BG869150	BG869150 602789028

ALIGNMENTS

RESULT 1
BG865470 747 bp mRNA EST 29-MAY-2001
LOCUS 602783596F1 NCI_CGAP_SG2 Mus musculus CDNA IMAGE:491062 5',
DEFINITION mRNA sequence.
ACCESSION BG865470.1 GI:14216010
VERSION BG865470
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 747)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Jeffrey B. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM10809 row: 0 column: 23
High quality sequence start: 2
High quality sequence stop: 746.
Location/Qualifiers
1..747
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:491062"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: salivary gland; Vector: pCMV-SPORT6; Site: 1; NCI; Site 2: Saliv; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.3 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library."

FEATURES

source

ORIGIN

Query Match 25.3%; Score 127; DB 4; Length 747;
Best Local Similarity 100.0%; Pred. No. 3.3e-27;
Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

JOURNAL
COMMENT

Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.jnl.gov
Plate: LLM10818 row: f column: 13
High quality sequence stop: 759.
Location/Qualifiers

FEATURES

source

1..767
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4913292"
/lab_host="DH10B (T1 phage-resistant)"
/clone_1lb="NCI CGAP SG2"
/note="Organ: salivary gland; Vector: PCMV-SPORT6; Site: 1;
NCI; Site_2: Saliv; Cloned unidirectionally. Primer: Oligo
dT. Average insert size 1.3 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

ORIGIN

Query Match

25.1%; Score 126; DB 4; Length 767;
Best Local Similarity 100.0%; Pred. No. 6.7e-27;
Matches 126; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 AAGATCTCTCTGTCGATCCCTTCTTCAAGAGTAGTAGTCACTAACTGA 350
Db 1 AAGATCTCTCTGTCGATCCCTTCTTCAAGAGTAGTAGTCACTAACTGA 60
QY 351 GTTCCATCTGAGAGTCACAGAGGCTTTGGGCAAGAGGAGGACCTGTT 410
Db 61 GTTCCATCTGAGAGTCACAGAGGCTTTGGGCAAGAGGAGGACCTGTT 120
QY 411 GTGTTG 416
Db 121 GTGTTG 126

RESULT 5

BM171830

LOCUS

BM171830 680 bp mRNA linear EST 04-DEC-2001
IMAGE:4922760 5', mRNA sequence.

DEFINITION

IMAGE:4922760 5', mRNA sequence.

BM171830

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BM171830 680 bp mRNA linear EST 04-DEC-2001
IMAGE:4922760 5', mRNA sequence.
BM171830 680 bp mRNA linear EST 04-DEC-2001
IMAGE:4922760 5', mRNA sequence.
EST. Mus musculus (house mouse)
Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 680)
Kale, P.I., Harsch, T.J., Folta, P.A., Nelson, D.O., Sanders, C.G. and
Prange, C.K.
The I.M.A.G.E. Consortium quality control effort: clone
resequencing for verification
Unpublished (2001)
Contact: Prange CK
The I.M.A.G.E. Consortium
Lawrence Livermore National Laboratory
Livermore, CA, USA
Email: help@image.jnl.gov
This read has been produced as part of the I.M.A.G.E. Consortium
quality control effort. High quality sequence is defined as having
100 or more base pairs with a phred quality value of 20 or greater,
where a sliding window of 4 base pairs with a phred quality value
of 15 or greater marks the beginning and end of the sequence. For
information on obtaining this clone, please contact

FEATURES

source

info@image.jnl.gov.
Plate: LLM10843 row: a column: 1
Seq primer: m3rp1
High quality sequence stop: 680.
Location/Qualifiers

1..680
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4922760"
/lab_host="DH10B (T1 phage-resistant)"
/clone_1lb="NCI CGAP SG2"
/note="Organ: salivary gland; Vector: PCMV-SPORT6; Site: 1;
NCI; Site_2: Saliv; Cloned unidirectionally. Primer: Oligo
dT. Average insert size 1.3 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

ORIGIN

Query Match

25.0%; Score 125; DB 4; Length 680;
Best Local Similarity 100.0%; Pred. No. 1.3e-26;
Matches 125; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 292 AAGATCTCTCTGTCGATCCCTTCTTCAAGAGTAGTAGTCACTAACTGAG 351
Db 1 AAGATCTCTCTGTCGATCCCTTCTTCAAGAGTAGTAGTCACTAACTGAG 60
QY 352 TTCCATCTGAGAGTCACAGAGGCTTTGGGCAAGAGGAGGACCTGTTG 411
Db 61 TTCCATCTGAGAGTCACAGAGGCTTTGGGCAAGAGGAGGACCTGTTG 120
QY 412 TGTTG 416
Db 121 TGTTG 125

RESULT 6

BG865845

LOCUS

BG865845 744 bp mRNA linear EST 29-MAY-2001
602788261F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4914158 5',
mRNA sequence.

DEFINITION

IMAGE:4914158 5', mRNA sequence.

BG865845

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BG865845 744 bp mRNA linear EST 29-MAY-2001
602788261F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4914158 5',
mRNA sequence.
EST. Mus musculus (house mouse)
Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 744)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.jnl.gov
Plate: LLM10820 row: j column: 15
High quality sequence stop: 708.
Location/Qualifiers

FEATURES

source

1..744
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4914158"
/lab_host="DH10B (T1 phage-resistant)"
/clone_1lb="NCI CGAP SG2"
/note="Organ: salivary gland; Vector: PCMV-SPORT6; Site: 1;

ORIGIN

Note: Site_2: Salt; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.3 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library."

Query Match 25.0%; Score 125; DB 4; Length 744;
Best Local Similarity 100.0%; Pred. No. 1.3e-26;
Matches 125; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 292 AGAATCTCTCTGTCATCTCTTCAAGCAGTAGTAGTGACTCCACTAACTGAG 351
|||||
DB 1 AGAATCTCTCTGTCATCTCTTCAAGCAGTAGTAGTGACTCCACTAACTGAG 60
|||||
QY 352 TTCCCATCTGAGAGTCCACAGAGGCTTTGGGCGAAGAGCAGAGGAAAGCACTGTTTG 411
|||||
DB 61 TTCCCATCTGAGAGTCCACAGAGGCTTTGGGCGAAGAGCAGAGGAAAGCACTGTTTG 120
|||||
QY 412 TGTGG 416
|||||
DB 121 TGTGG 125

RESULT 7
BG867523 938 bp mRNA linear EST 29-MAY-2001
LOCUS 602788460F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4919847 5',
DEFINITION mRNA sequence.

ACCESSION BG867523
VERSION BG867523.1 GI:14218063
KEYWORDS EST.

SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 938)
NIH-MGC http://mgc.nci.nih.gov/

UNPUBLISHED (1999)
National Institutes of Health, Mammalian Gene Collection (MGC)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov

Plate: LLAM10835 row: 9 column: 16
High quality sequence stop: 812.

Location/Qualifiers

FEATURES

source
1..938
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4919847"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: salivary gland; Vector: pCMV-Sport6; Site: 1;
Notr: Site 2: Salt; Cloned unidirectionally. Primer: Oligo
dT. Average insert size 1.3 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

ORIGIN

Query Match 25.0%; Score 125; DB 4; Length 938;
Best Local Similarity 100.0%; Pred. No. 1.4e-26;
Matches 125; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 292 AGAATCTCTCTGTCATCTCTTCAAGCAGTAGTAGTGACTCCACTAACTGAG 351
|||||
DB 1 AGAATCTCTCTGTCATCTCTTCAAGCAGTAGTAGTGACTCCACTAACTGAG 60
|||||
QY 352 TTCCCATCTGAGAGTCCACAGAGGCTTTGGGCGAAGAGCAGAGGAAAGCACTGTTTG 411
|||||

DB 61 TTCCCATCTGAGAGTCCACAGAGGCTTTGGGCGAAGAGCAGAGGAAAGCACTGTTTG 120
|||||
QY 412 TGTGG 416
|||||
DB 121 TGTGG 125

RESULT 8
BG869852 832 bp mRNA linear EST 29-MAY-2001
LOCUS 602789378F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4920739 5',
DEFINITION mRNA sequence.

ACCESSION BG869852
VERSION BG869852.1 GI:14220392
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 832)
NIH-MGC http://mgc.nci.nih.gov/

UNPUBLISHED (1999)
National Institutes of Health, Mammalian Gene Collection (MGC)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov

Plate: LLAM10837 row: 1 column: 20
High quality sequence stop: 807.

FEATURES

source
1..832
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4920739"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI CGAP SG2"
/note="Organ: salivary gland; Vector: pCMV-Sport6; Site: 1;
Notr: Site 2: Salt; Cloned unidirectionally. Primer: Oligo
dT. Average insert size 1.3 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

ORIGIN

Query Match 24.8%; Score 124; DB 4; Length 832;
Best Local Similarity 100.0%; Pred. No. 2.8e-26;
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 293 GGAATCTCTCTGTCATCTCTTCAAGCAGTAGTAGTGAAGTCACTAACTGAGT 352
|||||
DB 11 GGAATCTCTCTGTCATCTCTTCAAGCAGTAGTAGTGAAGTCACTAACTGAGT 70
|||||
QY 353 TTCCCATCTGAGAGTCCACAGAGGCTTTGGGCGAAGAGCAGAGGAAAGCACTGTTTG 412
|||||
DB 71 TTCCCATCTGAGAGTCCACAGAGGCTTTGGGCGAAGAGCAGAGGAAAGCACTGTTTG 130
|||||
QY 413 GTTGG 416
|||||
DB 131 GTTGG 134

RESULT 9
BG867502 1007 bp mRNA linear EST 29-MAY-2001
LOCUS 602788460F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4919935 5',
DEFINITION mRNA sequence.
ACCESSION BG867502

VERSION BG67502.1 GI:14218042
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 1007)
 NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished (1999)
 JOURNAL
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 Tissue Procurement: Jeffrey E. Green, M.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: LLAM10835 row: k column: 08
 High quality sequence stop: 801.
 Location/Qualifiers

FEATURES
 source
 1..1007
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:491935"
 /lab_host="DH10B (T1 phage-resistant)"
 /clone_1lb="NCI CGAP SG2"
 /note="Organ: salivary gland; Vector: pCMV-SPORT6; Site: 1; NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.3 kb. Constructed by Life Technologies. Note: this is a NCI CGAP Library."

ORIGIN

Query Match 24.4%; Score 122; DB 4; Length 1007;
 Best Local Similarity 100.0%; Pred. No. 1.2e-25;
 Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 295 ATCTCTCGTGCATCTCTTCAAGCAGTAGTAGTCCACTAACTGAGTTC 354
 1 ATCTCTCGTGCATCTCTTCAAGCAGTAGTAGTCCACTAACTGAGTTC 60
 DB 355 CCATCTGAGATCCACAGAGCGCTTTGGGCAAGACAGAGGAGGACCTGTTGT 414
 61 CCATCTGAGATCCACAGAGCGCTTTGGGCAAGACAGAGGAGGACCTGTTGT 120
 QY 415 TG 416
 DB 121 TG 122

RESULT 10
 BG873122 754 bp mRNA linear EST 29-MAY-2001
 LOCUS 602794520P1 NCI CGAP_SG2 Mus musculus cDNA clone IMAGE:4925621 5',
 DEFINITION mRNA sequence.
 ACCESSION BG873122
 VERSION BG873122.1 GI:14223662
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 754)
 NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished (1999)
 JOURNAL
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: LLAM10850 row: h column: 06
 High quality sequence stop: 754.
 Location/Qualifiers

ORIGIN

Query Match 24.0%; Score 120; DB 4; Length 754;
 Best Local Similarity 100.0%; Pred. No. 4.5e-25;
 Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 297 CTCTCTCGTGCATCTCTTCAAGCAGTAGTAGTCCACTAACTGAGTTC 356
 13 CTCTCTCGTGCATCTCTTCAAGCAGTAGTAGTCCACTAACTGAGTTC 72
 DB 357 ATCTGAGATCCACAGAGCGCTTTGGGCAAGACAGAGGAGGACCTGTTGT 416
 73 ATCTGAGATCCACAGAGCGCTTTGGGCAAGACAGAGGAGGACCTGTTGT 112

RESULT 11
 AA562244 249 bp mRNA linear EST 18-AUG-1997
 LOCUS V134B05.r1 Striatum mouse skin (#937311) Mus musculus cDNA clone
 DEFINITION IMAGE:974097 5' similar to gb:X01697 Mouse mRNA fragment for parotid secretory protein (MOUSE); mRNA sequence.

ACCESSION AA562244
 VERSION AA562244.1 GI:2333709
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

REFERENCE
 AUTHORS Merra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Scheinberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.
 TITLE The WashU-HMI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT Contact: Merra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.edu
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:554825

FEATURES
 source
 1..249
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"

/db_xref="taxon:10090"
 /clone="IMAGE:974097"
 /sex="females"
 /tissue_type="whole skin"
 /dev_stage="11 weeks old"
 /lab_host="GOLR (kanamycin resistant)"
 /clone_11b="Stratagene mouse skin (#937313)"
 /note="Organ: skin; Vector: pBluescript SK-; Site: 1:
 EcORI; Site: 2: XhoI; Cloned unidirectionally. Primer:
 01igo dr. Whole skin from 11 week old C57BL/6 female mice.
 Average insert size: 1.0 kb; Uni-ZAP XR Vector; ~5'
 adaptor sequence: 5' GAATTCGGCAGAG 3' ~3' adaptor
 sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3' "

ORIGIN

Query Match 23.8%; Score 119; DB 1; Length 249;
 Best Local Similarity 100.0%; Pred. No. 6.7e-25;
 Matches 119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 TCTCTGCTGTCATCTTTTCAAGACGATGAGTACTCTCACTAACTGATGCCA 357
 |||||
 DB 1 TCTCTGCTGTCATCTTTTCAAGACGATGAGTACTCTCACTAACTGATGCCA 60
 |||||
 QY 358 TCTGAGATCCACAGAGGCTTTGGGGCAAGAGCAGAGGACCTGTTGTGTTG 416
 |||||
 DB 61 TCTGAGATCCACAGAGGCTTTGGGGCAAGAGCAGAGGACCTGTTGTGTTG 119
 |||||

RESULT 12

CB575686

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

871 bp mRNA linear EST 02-APR-2003
 CB575686
 IMAGE:30292464 5', mRNA sequence.
 CB575686
 CB575686.1 GI:29495216
 EST.
 Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 871)
 NIH-MGC <http://mgi.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Dr. David Rowe
 CDNA Library Preparation: Invitrogen Corp
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: NDMA36 row: 9 column: 01
 High quality sequence start: 30
 High quality sequence stop: 528.
 Location/Qualifiers

FEATURES

source

1..871
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:30292464"
 /tissue_type="embryonic limb, maxilla and mandible"
 /lab_host="DH10B (phage-resistant)"
 /note="Vector: PCMV-SPORT6.1; Site 1: EcoRV; Site 2: NotI;
 Normalized, full-length enriched library from pool of
 mouse embryonic limb, maxilla and mandible, embryonic day
 17.5, 18.5 and newborn (mandible (5, 4 and 1 limb and jaw
 equivalents from respective days). Cloned directionally,
 oligo-dT primed (5'-GACTAGCTTAGATCGAGCGGCCGCCCTT)15-3'.
 Size selected for the >1kb fragments, average insert size
 1.2 kb. Normalization to Cot 7.5. Tissue contributed by

ORIGIN

Query Match 23.5%; Score 117.6; DB 6; Length 871;
 Best Local Similarity 96.8%; Pred. No. 2.6e-24;
 Matches 120; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 223 GAGTCTCTCTGTCATCTTTTCAAGACGATGAGTACTCTCACTAACTGAGT 352
 |||||
 DB 6 GAGTCTCTCTGTCATCTTTTCAAGACGATGAGTACTCTCACTAACTGAGT 65
 |||||
 QY 353 TCCCATCTGAGATCCACAGAGGCTTTGGGGCAAGAGCAGAGGACCTGTTGT 412
 |||||
 DB 66 TCCCATCTGAGATCCACAGAGGCTTTGGGGCAAGAGCAGAGGACCTGTTGT 125
 |||||
 QY 413 GTTG 416
 |||||
 DB 126 GTTG 129

RESULT 13

AO974026

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

563 bp DNA linear GSS 28-JAN-2000
 AO974026
 RPCI-23-331F18.TV RPCI-23 Mus musculus genomic clone
 RPCI-23-331F18, genomic survey sequence.
 AO974026
 AO974026.1 GI:6804479
 GSS.
 Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 563)
 Zhao, S., Nierman, W., Feldblum, T., Malek, J., Shatman, S.,
 Akintef, B., Levins, M., McGinn, S., Tsegaye, G., Geer, K., Krol, M., de
 Jong, P. and Fraser, C.M.
 Mouse BAC End Sequences from Library RPCI-23
 Unpublished (1999)
 Other_GSSs: RPCI-23-331F18.TV
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@igf.org
 Clones are derived from the mouse BAC library RPCI-23. For BAC
 library availability, please contact Pieter de Jong
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)
 or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tlgr.org/cdb/bac_ends/mouse/bac_end_intro.html
 Plate: 331 row: F column: 18
 Seq primer: SP6
 Class: BAC ends.
 Location/Qualifiers

FEATURES

source

1..563
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-23-331F18"
 /sex="Female"
 /lab_host="DH10B"
 /clone_11b="RPCI-23"
 /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site: 1:
 EcoRI; Site: 2: EcoRI; Female C57BL/6J mouse kidney and/or
 brain genomic DNA was isolated and partially digested
 with a combination of EcoRI and EcoRI Methylase. Size
 selected DNA was cloned into the pBACe3.6 vector at the
 EcoRI sites. The ligation products were transformed into
 DH10B electrocompetent cells (BRL Life Technologies)."

Fri Feb 11 08:00:44 2005

us-09-926-375b-7_copy_11500_12000.rst

Job time : 2175.78 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 20:12:02 ; Search time 341.632 Seconds
(without alignments)
8681.254 Million cell updates/sec

Title: US-09-926-375B-7_COPY_11500_12000

Perfect score: 501
Sequence: 1 aaccacacagacatcttc.....gactcatcgcgggttcgttt 501

Scoring table: IDENTITY NUC
Gapop 10_0, Gapext 1_0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	501	100.0	17732	3 AAC68300	Aac68300 Lama2/APP
2	501	100.0	20623	3 AAC68294	Aac68294 Lama2/APP
3	40.8	8.1	7524	4 AAK89798	Aak89798 Human dig
4	37	7.4	3555	8 ACA28334	ACA28334 Prokaryot
5	36.8	7.3	48088	11 ACN43856	ACN43856 Mouse gen
6	34.8	6.9	1470	3 AAC68112	Aac68112 Human sec
7	34.6	6.9	4352	13 ADQ39113	Adq39113 Human SNP
8	34.4	6.9	110000	13 ABD32625_4	Continuation (5 of
9	34.2	6.8	7991	13 ADT05514	Adt05514 Haemophil
10	34	6.8	58651	12 ADQ97152	Adq97152 Human can
11	33.8	6.7	23473	12 ADQ97379	Adq97379 Human can
12	33.6	6.7	295096	11 ACN44068	ACN44068 Mouse gen
13	33.6	6.6	1089	3 AAC51458	Aac51458 Arabidops
14	33.2	6.6	1417	3 AAC51477	Aac51477 Arabidops
15	33.2	6.6	1431	3 AAC46837	Aac46837 Arabidops
16	33.2	6.6	5656	4 ABL03629	ABl03629 Drosophil
17	33.2	6.6	8477	4 ABL03628	ABl03628 Drosophil
18	32.8	6.5	2000	8 ADA71938	Ada71938 Rice gene
19	32.6	6.5	311	6 ABR78273	ABr78273 Human ORF
20	32.6	6.5	89829	13 ABD33301	ABd33301 Murine ca

C	21	32.6	6.5	110000	12 ADO34927_0	Ado34927 Human vol
C	22	32.6	6.5	197526	13 ABD33389	ABd33389 Human can
C	23	32.4	6.5	2000	10 ADK62379	ADk62379 Disease t
C	24	32.4	6.5	2508	10 ADC70455	ADc70455 Yeast 250
C	25	32.4	6.5	83043	10 ADF77343_19	Continuation (20 o
C	26	32.2	6.4	37925	4 AAK72372	Aak72372 Human imm
C	27	32.2	6.4	37925	12 ADJ12847	Adj12847 DNA fragm
C	28	32.2	6.4	211257	11 ACN44200	ACn44200 Mouse gen
C	29	32.2	6.4	1808	13 ADS57667	ADs57667 Bacterial
C	30	32	6.4	76138	12 ADQ97334	Adq97334 Mouse can
C	31	32	6.4	155083	12 ADQ97891	Adq97891 Mouse can
C	32	31.8	6.3	825	8 ACA28877	ACA28877 Prokaryot
C	33	31.8	6.3	2054	3 AAA64217	Aaa64217 Fast skel
C	34	31.8	6.3	5969	6 ABL992326	ABl992326 Chemical
C	35	31.8	6.3	5969	8 ADN20380	ADa20380 Prostate
C	36	31.8	6.3	5969	8 ADA84187	Ada84187 Human ren
C	37	31.8	6.3	23169	5 ABA15819	ABa15819 Human ner
C	38	31.8	6.3	23172	5 ABA15818	ABa15818 Human ner
C	39	31.8	6.3	100608	13 ABD32769	ABd32769 Human can
C	40	31.8	6.3	322101	10 AAD58431	AAd58431 Human can
C	41	31.8	6.3	322101	12 ADJ35046	ADj35046 Human pro
C	42	31.6	6.3	1023	13 ADS59093	ADs59093 Bacterial
C	43	31.6	6.3	10155	10 ABQ83443	ABq83443 Papaya le
C	44	31.4	6.3	307	8 ABZ25357	ABz25357 Murine in
C	45	31.4	6.3	669	9 ACL23453	ACl23453 DNA clone

ALIGNMENTS

RESULT 1	AAC68300	standard; DNA; 17732 BP.
ID	AAC68300	
XX		
AC	AAC68300;	
XX		
DT	15-SEP-2003 (revised)	
DT	20-FEB-2001 (first entry)	
XX		
DE	Lama2/APPa plasmid coding sequence.	
XX		
KW	Transgenic animal; salivary protein; phytase; phosphorus; animal growth;	
KW	environmental pollution; pig; ds.	
XX		
OS	Mus musculus.	
OS	Escherichia coli.	
OS	Chimeric.	
XX		
PN	WO200064247-A1.	
XX		
PD	02-NOV-2000.	
XX		
PF	20-APR-2000; 2000WO-CA000430.	
XX		
PR	23-APR-1999; 99US-0130508P.	
XX		
PA	(UYGU-) UNIV GUELPH.	
XX		
PI	Forsberg CW, Golovan S, Phillips JP;	
XX		
DR	WPI, 2000-687245/67.	
DR	P-PSDB; AAB36263.	
XX		
PT	Transgenic non-human animal for gastrointestinal tract specific	
PT	expression of a protein, preferably phytase, comprises a nucleic acid	
PT	sequence including a heterologous transgene construct encoding the	
PT	protein.	
XX		
XX	Claim 14; Fig 23; 152p; English.	
CC	The present invention provides transgenic animals which produce desired	
CC	proteins, in this case pigs which expresses phytase in the salivary	
CC	gland. Low phytase production levels result in phytase in the diet being	

CC excreted and causing phosphorus contamination in water, as well as
 CC reducing the growth of animals. The invention provides a number of
 CC transgenes containing the E. coli APPA phytase coding sequence. (Updated
 CC on 15-SEP-2003 to standardise OS field)

XX Sequence 17732 BP; 4719 A; 4125 C; 4168 G; 4719 T; 0 U; 1 Other;

Query Match 100.0%; Score 501; DB 3; Length 17732;
 Best Local Similarity 100.0%; Pred. No. 4e-151;
 Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACCCACAGAACATCTTCCATGATTCGATTCCTTCTAGACAAATCTGAA 60
 DB AACCCACAGAACATCTTCCATGATTCGATTCCTTCTAGACAAATCTGAA 11559
 QY CTGACCCCTTGTTGGAGTCTCAAGTTTGACAATAGGGCTTTGAACCTGGACAAGGT 120
 DB CTGACCCCTTGTTGGAGTCTCAAGTTTGACAATAGGGCTTTGAACCTGGACAAGGT 11619
 QY 121 CCATCACTGTACCCCAAGCATCTGGGTGACCTTTGGGTTGGAATCTTGGCTAACCTT 180
 DB 11620 CCATCACTGTACCCCAAGCATCTGGGTGACCTTTGGGTTGGAATCTTGGCTAACCTT 11679
 QY 181 AGATATTTCTTGGAGTATCTTTAGAACATCCAGGAATAGGGCTTGAATCTCATCTG 240
 DB 11680 AGATATTTCTTGGAGTATCTTTAGAACATCCAGGAATAGGGCTTGAATCTCATCTG 11739
 QY 241 GGACCACAATATATAGTCAACCTTAGAATCCAGAGATCTGTGACAGAAACAAGATCTCT 300
 DB 11740 GGACCACAATATATAGTCAACCTTAGAATCCAGAGATCTGTGACAGAAACAAGATCTCT 11799
 QY 301 CTGTGTGATCTCTTCTTCAAGAGATGATGATCTCACTAACTGATCTCCATCT 360
 DB 11800 CTGTGTGATCTCTTCTTCAAGAGATGATGATCTCACTAACTGATCTCCATCT 11859
 QY 361 GAGAGTCCACAGAGGCTTTGGGGCAAGAGAGAGGAAAGGCACTGTTTGGTAA 420
 DB 11860 GAGAGTCCACAGAGGCTTTGGGGCAAGAGAGAGGAAAGGCACTGTTTGGTAA 11919
 QY 421 AGTTTGAATCTTAACAATTTGAAGACATAGATGATCTGTGCACTTAACAACCT 480
 DB 11920 AGTTTGAATCTTAACAATTTGAAGACATAGATGATCTGTGCACTTAACAACCT 11979
 QY 481 AGACTCATGTGGGTTCTGTTT 501
 DB 11980 AGACTCATGTGGGTTCTGTTT 12000

RESULT 2

AAC68294 standard; DNA; 20623 BP.

XX AAC68294;
 AC 15-SEP-2003 (revised)
 DT 20-FEB-2001 (first entry)
 XX
 DE Lama2/APPA plasmid coding sequence.
 XX
 KM Transgenic animal; salivary protein; phytase; phosphorus; animal growth;
 KM environmental pollution; pig; ds.
 OS Mus musculus.
 OS Escherichia coli.
 OS Chimeric.
 XX
 XX WO20064247-A1.
 XX 02-NOV-2000.
 XX 20-APR-2000; 2000WO-CA000430.
 XX 23-APR-1999; 99US-0130508P.
 XX

XX (UYGU-) UNIV GUELPH.
 PA Forsberg CW, Golovan S, Phillips JP;
 PI MPI; 2000-687245/67.
 DR P-PSDB; AAB36257.
 XX

PT Transgenic non-human animal for gastrointestinal tract specific
 PT expression of a protein, preferably phytase, comprises a nucleic acid
 PT sequence including a heterologous transgene construct encoding the
 PT protein.
 XX

PS Claim 56; Fig 5; 152pp; English.

CC The present invention provides transgenic animals which produce desired
 CC proteins, in this case pigs which express phytase in the salivary
 CC gland. Low phytase production levels result in phytate in the diet being
 CC excreted and causing phosphorus contamination in water, as well as
 CC reducing the growth of animals. The invention provides a number of
 CC transgenes containing the E. coli APPA phytase coding sequence. (Updated
 CC on 15-SEP-2003 to standardise OS field)

XX Sequence 20623 BP; 5449 A; 4847 C; 4902 G; 5424 T; 0 U; 1 Other;

Query Match 100.0%; Score 501; DB 3; Length 20623;
 Best Local Similarity 100.0%; Pred. No. 4.3e-151;
 Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACCCACAGAACATCTTCCATGATTCGATTCCTTCTAGACAAATCTGAA 60
 DB 11500 AACCCACAGAACATCTTCCATGATTCGATTCCTTCTAGACAAATCTGAA 11559
 QY 61 CTGACCCCTTGTTGGAGTCTCAAGTTTGACAATAGGGCTTTGAACCTGGACAAGGT 120
 DB 11560 CTGACCCCTTGTTGGAGTCTCAAGTTTGACAATAGGGCTTTGAACCTGGACAAGGT 11619
 QY 121 CCATCACTGTACCCCAAGCATCTGGGTGACCTTTGGGTTGGAATCTTGGCTAACCTT 180
 DB 11620 CCATCACTGTACCCCAAGCATCTGGGTGACCTTTGGGTTGGAATCTTGGCTAACCTT 11679
 QY 181 AGATATTTCTTGGAGTATCTTTAGAACATCCAGGAATAGGGCTTGAATCTCATCTG 240
 DB 11680 AGATATTTCTTGGAGTATCTTTAGAACATCCAGGAATAGGGCTTGAATCTCATCTG 11739
 QY 241 GGACCACAATATATAGTCAACCTTAGAATCCAGAGATCTGTGACAGAAACAAGATCTCT 300
 DB 11740 GGACCACAATATATAGTCAACCTTAGAATCCAGAGATCTGTGACAGAAACAAGATCTCT 11799
 QY 301 CTGTGTGATCTCTTCTTCAAGAGATGATGATCTCACTAACTGATCTCCATCT 360
 DB 11800 CTGTGTGATCTCTTCTTCAAGAGATGATGATCTCACTAACTGATCTCCATCT 11859
 QY 361 GAGAGTCCACAGAGGCTTTGGGGCAAGAGAGGAAAGGCACTGTTTGGTAA 420
 DB 11860 GAGAGTCCACAGAGGCTTTGGGGCAAGAGAGGAAAGGCACTGTTTGGTAA 11919
 QY 421 AGTTTGAATCTTAACAATTTGAAGACATAGATGATCTGTGCACTTAACAACCT 480
 DB 11920 AGTTTGAATCTTAACAATTTGAAGACATAGATGATCTGTGCACTTAACAACCT 11979
 QY 481 AGACTCATGTGGGTTCTGTTT 501
 DB 11980 AGACTCATGTGGGTTCTGTTT 12000

RESULT 3

AAK89798 standard; DNA; 7524 BP.

XX AAK89798;
 AC 05-NOV-2001 (first entry)
 XX

XX Human digestive system antigen genomic sequence SEQ ID NO: 3374.
DE
XX
XX Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW digestive system disorder; Meckel's diverticulum; de.
XX
OS Homo sapiens.
XX
XX WO200155314-A2.
XX
XX 02-AUG-2001.
XX
XX
XX 17-JAN-2001; 2001WO-US001324.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180628P.
XX 24-FEB-2000; 2000US-0184664P.
XX 02-MAR-2000; 2000US-0186350P.
XX 16-MAR-2000; 2000US-0189874P.
XX 17-MAR-2000; 2000US-0190076P.
XX 18-APR-2000; 2000US-0198123P.
XX 19-MAY-2000; 2000US-0205515P.
XX 07-JUN-2000; 2000US-0209467P.
XX 28-JUN-2000; 2000US-0214866P.
XX 30-JUN-2000; 2000US-0215135P.
XX 07-JUL-2000; 2000US-0216647P.
XX 11-JUL-2000; 2000US-0217487P.
XX 14-JUL-2000; 2000US-0217496P.
XX 26-JUL-2000; 2000US-0220963P.
XX 26-JUL-2000; 2000US-0220964P.
XX 14-AUG-2000; 2000US-0224518P.
XX 14-AUG-2000; 2000US-0224519P.
XX 14-AUG-2000; 2000US-0225213P.
XX 14-AUG-2000; 2000US-0225214P.
XX 14-AUG-2000; 2000US-0225266P.
XX 14-AUG-2000; 2000US-0225267P.
XX 14-AUG-2000; 2000US-0225268P.
XX 14-AUG-2000; 2000US-0225270P.
XX 14-AUG-2000; 2000US-0225447P.
XX 14-AUG-2000; 2000US-0225477P.
XX 14-AUG-2000; 2000US-0225757P.
XX 14-AUG-2000; 2000US-0225758P.
XX 14-AUG-2000; 2000US-0225759P.
XX 18-AUG-2000; 2000US-0226279P.
XX 22-AUG-2000; 2000US-0226681P.
XX 22-AUG-2000; 2000US-0226868P.
XX 22-AUG-2000; 2000US-0227182P.
XX 23-AUG-2000; 2000US-0227009P.
XX 30-AUG-2000; 2000US-0228924P.
XX 01-SEP-2000; 2000US-0229287P.
XX 01-SEP-2000; 2000US-0229343P.
XX 01-SEP-2000; 2000US-0229344P.
XX 01-SEP-2000; 2000US-0229345P.
XX 05-SEP-2000; 2000US-0229509P.
XX 05-SEP-2000; 2000US-0229513P.
XX 06-SEP-2000; 2000US-0230437P.
XX 06-SEP-2000; 2000US-0230438P.
XX 08-SEP-2000; 2000US-0231242P.
XX 08-SEP-2000; 2000US-0231243P.
XX 08-SEP-2000; 2000US-0231244P.
XX 08-SEP-2000; 2000US-0231413P.
XX 08-SEP-2000; 2000US-0231414P.
XX 08-SEP-2000; 2000US-0232081P.
XX 08-SEP-2000; 2000US-0232081P.
XX 12-SEP-2000; 2000US-0231968P.
XX 14-SEP-2000; 2000US-0232397P.
XX 14-SEP-2000; 2000US-0232398P.
XX 14-SEP-2000; 2000US-0232399P.
XX 14-SEP-2000; 2000US-0232400P.
XX 14-SEP-2000; 2000US-0232401P.
XX 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235634P.
PR 27-SEP-2000; 2000US-0235636P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0244647P.
PR 08-NOV-2000; 2000US-02446475P.
PR 08-NOV-2000; 2000US-02446476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251797P.
PR 06-DEC-2000; 2000US-0251856P.
PR 06-DEC-2000; 2000US-0251868P.

PI	Mang L, Zamudio C, Malone C, Hasebeck R, Ohlsen KL, Zyskind JW;
FI	Wall D, Trivick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;
DR	WPI, 2003-029926/02.
XX	P-RSDB; ABU24464.
PT	New antisense nucleic acids, useful for identifying proteins or screening
PT	for homologous nucleic acids required for cellular proliferation to
PT	isolate candidate molecules for rational drug discovery programs.
XX	
XX	Claim 14; SEQ ID NO 16204; 1766pp; English.
CC	The invention relates to an isolated nucleic acid comprising any one of
CC	the 613 antisense sequences given in the specification where expression
CC	of the nucleic acid inhibits proliferation of a cell. Also included are:
CC	(1) a vector comprising a promoter operably linked to the nucleic acid
CC	encoding a polypeptide whose expression is inhibited by the antisense
CC	nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC	polypeptide or its fragment whose expression is inhibited by the
CC	antisense nucleic acid; (4) an antibody capable of specifically binding
CC	the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC	proliferation or the activity of a gene in an operon required for
CC	proliferation; (7) identifying a compound that influences the activity of
CC	the gene product or that has an activity against a biological pathway
CC	required for proliferation, or that inhibits cellular proliferation; (8)
CC	identifying a gene required for cellular proliferation or the biological
CC	pathway in which a proliferation-required gene or its gene product lies
CC	or a gene on which the test compound that inhibits proliferation of an
CC	organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC	compound's activity; (11) a culture comprising strains in which the gene
CC	product is overexpressed or underexpressed; (12) determining the extent
CC	to which each of the strains is present in a culture or collection of
CC	strains; or (13) identifying the target of a compound that inhibits the
CC	proliferation of an organism. The antisense nucleic acids are useful for
CC	identifying proteins or screening for homologous nucleic acids required
CC	for cellular proliferation to isolate candidate molecules for rational
CC	drug discovery programs, or for screening homologous nucleic acids
CC	required for proliferation in cells other than S. aureus, S. typhimurium,
CC	K. pneumoniae or P. aeruginosaes. The present sequence is one of the target
CC	prokaryotic essential genes. Note: The sequence data for this patent did
CC	not form part of the printed specification, but was obtained in
CC	electronic format directly from WIPO at
CC	ftp.wipo.int/pub/published_pct_sequences
XX	
SQ	Sequence 3555 BP; 1445 A; 358 C; 685 G; 1067 T; 0 U; 0 Other;
Query Match	7.4%; Score 37; DB 8; Length 3555;
Best Local Similarity	56.0%; Pred. No. 0.82;
Matches	70; Conservative 0; Mismatches 55; Indels 0; Gaps 0;
OY	181 AGATATTTCCTTGAGATATCTTTAAGAATCGAGAAATAGGGCTTGATTCTCATCCTG 240 Db
Db	1947 AGATTAATATACAGGAATAAATTATTAATATAGACGAGGCCCTTAGTGCATAATACCAAGATA 2006
OY	241 GGACCACAATTAAGTCACTCACTTAAGATCCCAGGAGATGTGCACAGAAACAAGATCTCT 300
Db	2007 TATAGCTAATAAAAATAAACCTTTAATATATGAATATCAACAGAAAAACTAAGAACTAT 2066
OY	301 CTCGT 305
Db	2067 ATGTG 2071
RESULT 5	
ACN43856/C	
ID ACN43856 standard; DNA; 49088 BP.	
AC ACN43856;	
XX	
DT 18-NOV-2004 (first entry)	
XX	
DE Mouse genomic sequence MCGI0132.	

XX Cytostatic; carcinoma; lymphoma; cancer; murine; gene; ss.
XX Mus musculus.
XX WO2003073826-A2.
XX 12-SEP-2003.
XX 28-FEB-2003; 2003WO-US006235.
XX 01-MAR-2002; 2002US-00087192.
XX (SAGR-) SAGRES DISCOVERY.
XX Morris DW;
XX WPI; 2003-328604/31.
XX Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
XX comprises a nucleotide sequence.
XX
XX Claim 1; SEQ ID NO 13; Opp; English.
XX The present invention relates to novel DNA and protein sequences which
XX are associated with carcinomas. The sequences are useful for: (i) for
XX screening drug candidates; (ii) for screening of bioactive agent capable
XX of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
XX a bioactive agent capable of modulating the activity of CAP; (iv) for
XX evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
XX carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
XX carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a block; (x)
XX for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
XX determining Carcinoma Associated (CA) gene copy number. In addition, the
XX CA genes are useful as DNA vaccines and the CAP are useful as markers of
XX carcinoma including lymphoma. The present sequence is one such CA coding
XX sequence. Note: This patent is an equivalent to basic patent
XX US2002182586A1, for which no sequence data was published
XX
XX Sequence 49088 BP; 13195 A; 9519 C; 10290 G; 14938 T; 0 U; 1146 Other;
XX
XX Query Match 7.3%; Score 36.8; DB 11; Length 49088;
XX Best Local Similarity 59.6%; Pred. No. 3.2;
XX Matches 62; Conservative 0; Mismatches 42; Indels 0; Gaps 0;
XX
XX QY 255 GTCACCTAGATCCGAGAGATGTCGAGAAACAAGATCTCTCGTGATCCT 314
XX 45517 GTCTCAAAACAACTGCGAGAGTGCACCACTGAAATTAAGTTTAAACCCCTGAACT 45458
XX
XX Db 315 TCTTCAAGCAGTAGTAGTCACTCACTAACTGAGTTCCCAT 358
XX 45457 ACATTAAGCAAGAAAGAAATGACCCCACTAAGTAGTCTTTCAT 45414
XX
XX Db 45457 ACATTAAGCAAGAAAGAAATGACCCCACTAAGTAGTCTTTCAT 45414
XX
XX RESULT 6
XX AAC68112
XX ID AAC68112 standard; cDNA; 1470 BP.
XX
XX AAC68112;
XX
XX 20-FEB-2001 (first entry)
XX
XX Human secreted protein cDNA sequence #32.
XX
XX Cytostatic; immunosuppressive; neurotropic; neuroprotective; antiviral;
XX antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
XX vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX cardiact; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX neurological disease; infection; human; secreted protein; ss.
XX
XX Homo sapiens.
XX
XX WO200058335-A1.

XX 05-OCT-2000.
XX 22-MAR-2000; 2000WO-US007534.
XX 26-MAR-1999; 99US-0126598P.
XX 22-DEC-1999; 99US-0171504P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.
XX Rosen CA, Ruben SM, Komatsoulis G;
XX WPI; 2000-611702/58.
XX P-PSDB; AAB37379.
XX
XX Nucleic acids encoding human secreted proteins, used to treat, prevent,
XX ameliorate or diagnose conditions such as cancer, and autoimmune diseases
XX e.g. arthritis.
XX
XX Claim 1; Page 337; 387pp; English.
XX
XX The invention relates to the isolation of genes AAC68081-C68127 encoding
XX 47 human secreted proteins AAB37348-B37394. The genes can be used to
XX generate fusion proteins by linking to the gene for the human
XX immunoglobulin G Fc portion (AAC68072) for increasing the stability of
XX the fusion protein as compared to the human protein only. The genes and
XX proteins are useful for preventing, ameliorating or treating medical
XX conditions, e.g. by protein or gene therapy. The genes are isolated from
XX a range of human tissues disclosed in the specification. The nucleic
XX acids, proteins, antibodies and (ant)agonists are useful in the
XX diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
XX ovarian cancer, and other cancers of the adrenal gland, bone, bone
XX marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b)
XX immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic
XX anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease;
XX multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
XX cardiovascular disorders such as myocardial ischaemia; (d) wound healing
XX ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
XX infectious diseases such as viral, bacterial, fungal and parasitic
XX infections
XX
XX Sequence 1470 BP; 386 A; 342 C; 355 G; 382 T; 0 U; 5 Other;
XX
XX Query Match 6.9%; Score 34.8; DB 3; Length 1470;
XX Best Local Similarity 54.8%; Pred. No. 2.8;
XX Matches 69; Conservative 0; Mismatches 57; Indels 0; Gaps 0;
XX
XX QY 373 GAGGCTTTGGGCGAAGAGAGAGGAGGCACTGTTGTGTGTAAGTTTGAATCT 432
XX 71 GAAGCAGTGGGGCAAGAGAGAGAGAGGAGTGTGCTGATGTTTAAAGCTTGACAGT 130
XX
XX QY 433 AACAAATTTGAGACATAGATGACATTTGTGTGACACTAACAACACTAGACTGATGG 492
XX 131 CACATATAGATTTCCTCCAGGTACCCCTGCTGCCCCCAAGACATCCAAAGCAAGATGG 190
XX
XX Db 493 GTTCTG 498
XX 191 ATGGTG 196
XX
XX Db 191 ATGGTG 196
XX
XX RESULT 7
XX ADQ39113/C
XX ID ADQ39113 standard; DNA; 4352 BP.
XX
XX ADQ39113;
XX
XX 18-NOV-2004 (first entry)
XX
XX Human SNP containing myocardial infarction-associated gene, SEQ ID 776.
XX
XX Myocardial infarction; detection; single nucleotide polymorphism; SNP;
XX cardiact; gene therapy; human; gene; de.

Query Match 6.8%; Score 34.2; DB 13; Length 7991;
Best Local Similarity 53.3%; Pred. No. 9.7;
Matches 72; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 216 GAAATAGGAGCTGATCTCATCTGAGACACATATAGTCACCTAGATCCAGAG 275
DB 2844 GCATTATATTTTCTTTTGTGATGTAATTCACAAACCTTATCCATCCCGGG 2785
QY 276 ATCTGTCAGAGAAACAAAGATCTCTCTGTCGATCTTTCTTCAAGCAGTAGTGTG 335
DB 2784 ATGCAGAAATATAAATGGATATTGCTGCTTACCTTTTACTTAACACATTTATTG 2725
QY 336 ACTCCACTAACTGA 350
DB 2724 GTTTTACGCTATTTA 2710

RESULT 10
ADQ97152/c
ID ADQ97152 standard; DNA; 58651 BP.
AC ADQ97152;
XX 07-OCT-2004 (first entry)
XX Human cancer associated sequence HD08-008, SEQ ID 128.
XX De
XX Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Human; ds.
XX KM
XX Homo sapiens.
XX OS
XX WO2004060304-A2.
XX PN
XX 22-JUL-2004.
XX PD
XX 22-DEC-2003; 2003WO-US041389.
XX PF
XX 27-DEC-2002; 2002US-00330773.
XX PR
XX (SAGR-) SAGRES DISCOVERY INC.
XX PA
XX Morris DW, Malandro MS;
XX PI
XX WPI; 2004-543781/52.
XX DR
XX New isolated cancer associated nucleic acids comprising at least 10
XX PT contiguous nucleotides, useful for diagnosing, preventing and/or treating
XX PT cancers such as leukemia and lymphoma.
XX PS
XX Claim 1; SEQ ID NO 128; 199pp; English.
XX CC The present invention relates to cancer associated sequences (ADQ97025-
XX CC ADQ98004). The sequences are useful for the diagnosis, prevention and/or
XX CC treatment of cancer, such as leukemia and lymphoma. Note: The sequence
XX CC data for this patent did not form part of the printed specification, but
XX CC was obtained in electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences.

QY Sequence 58651 BP; 15392 A; 12241 C; 13719 G; 17299 T; 0 U; 0 Other;
Query Match 6.8%; Score 34; DB 12; Length 58651;
Best Local Similarity 54.9%; Pred. No. 29;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 380 TTGGGCAAGACGAGGAAAGCACTGTTGTGTGTAAGTTTGACTTAACAAT 439
DB 58574 TGCAGCTTAAGTGAAGGCGGTGGCACTGAGGTGAGAGGAGATGGGCACTGACAGA 58515
QY 440 TTGAAGACATGATGATGATGTCTCAAGCTAACAACAACCTTGAATGAGGTTCTGT 499
DB 58514 GAGTGAACAGAGCAGCAGCTAGGCTATCTAGAAACAAACAGTACATCAGCGATCAGT 58455
QY 500 TT 501

DB 58454 TT 58453

RESULT 11
ADQ97379/c
ID ADQ97379 standard; DNA; 23473 BP.
XX ADQ97379;
XX AC
XX 07-OCT-2004 (first entry)
XX DT
XX Human cancer associated sequence HD08-040, SEQ ID 356.
XX DE
XX Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Human; ds.
XX KM
XX Homo sapiens.
XX OS
XX WO2004060304-A2.
XX PN
XX 22-JUL-2004.
XX PD
XX 22-DEC-2003; 2003WO-US041389.
XX PF
XX 27-DEC-2002; 2002US-00330773.
XX PR
XX (SAGR-) SAGRES DISCOVERY INC.
XX PA
XX Morris DW, Malandro MS;
XX PI
XX WPI; 2004-543781/52.
XX DR
XX New isolated cancer associated nucleic acids comprising at least 10
XX PT contiguous nucleotides, useful for diagnosing, preventing and/or treating
XX PT cancers such as leukemia and lymphoma.
XX PS
XX Claim 1; SEQ ID NO 356; 199pp; English.
XX CC The present invention relates to cancer associated sequences (ADQ97025-
XX CC ADQ98004). The sequences are useful for the diagnosis, prevention and/or
XX CC treatment of cancer, such as leukemia and lymphoma. Note: The sequence
XX CC data for this patent did not form part of the printed specification, but
XX CC was obtained in electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences.

QY Sequence 23473 BP; 5653 A; 5992 C; 6186 G; 5642 T; 0 U; 0 Other;
Query Match 6.7%; Score 33.8; DB 12; Length 23473;
Best Local Similarity 48.2%; Pred. No. 22;
Matches 95; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 283 AGAGAAACAGAGATCTCTCTGTCGTCATCTTTCTTCAAGCAGTAGTAGTCCAC 342
DB 22832 ATATATATATATATATATGTCCTTCAAGTCTTTCTTCAACACAGCATCCTGTCTGTG 22773
QY 343 TAAACTGAGTTCCTCATCTGAGATCCACAGAGAGCTTTGGGCAAGAGAGGAGG 402
DB 22772 ACCTAGAGAGTCAATCTTGCACCTTCACAGACTGCTGTGGCTGGGAGAGGAGG 22713
QY 403 CACTGTTGTGTGTTGTTGAAGTTTGAATTTGAACATATGATGATGATG 462
DB 22712 GAATGAGGCTATTGTGTAAGAGAGCTTGAACCGCTGTGGATAGCTGCTCATATGTC 22653
QY 463 TCAGACTTAACAACACC 479
DB 22652 ACTGCCCACTTCTTC 22636

RESULT 12
ACN44068/c
ID ACN44068 standard; DNA; 295096 BP.
XX ACN44068;
XX AC

PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-014187P.
PR 01-JUL-1999; 99US-0141842P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142820P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145591P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 03-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147126P.
PR 06-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.

PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 6.6%; Score 33.2; DB 3; Length 1089;
Best Local Similarity 61.6%; Pred. No. 8.1;
Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 337 CTCACCTAACTGATCCATCTGAGAGTCACAGAGCTTTGGGACAGAGCAG 396
DB 220 CTTACACACCTTACGACCTTTGTGAGGACACAGAGCGTTTCCCAATAGCAG 161
QY 397 GGAAGGACCTGTTGTGTTGTAAG 422
DB 160 AGGGTCACGTTTGGCGTTGTAGAG 135

RESULT 14
AAC51477/c
ID AAC51477 standard; DNA; 1417 BP.

AC AAC51477;
XX 18-OCT-2000 (first entry)
DT XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 68668.
DE XX
XX

KM Hybridisation assay; genetic mapping; gene expression control;
KM protein identification; signal transduction pathway; metabolic pathway;
KM promoter; termination sequence; ss.

OS Arabidopsis thaliana.

PN EP1033405-A2.

XX 06-SEP-2000.

PD 25-FEB-2000; 2000EP-00301439.

XX 25-FEB-1999; 99US-0121825P.

XX 05-MAR-1999; 99US-0123180P.

PR 09-MAR-1999; 99US-0125548P.

PR 23-MAR-1999; 99US-0126264P.

PR 25-MAR-1999; 99US-0126785P.

PR 01-APR-1999; 99US-0127462P.

PR 06-APR-1999; 99US-0128234P.

PR 08-APR-1999; 99US-0128714P.

PR 16-APR-1999; 99US-0129645P.

PR 19-APR-1999; 99US-0130077P.

PR 21-APR-1999; 99US-0130510P.

PR 23-APR-1999; 99US-0130891P.

PR 28-APR-1999; 99US-0131449P.

PR 30-APR-1999; 99US-0132048P.

PR 04-MAY-1999; 99US-0132407P.

PR 05-MAY-1999; 99US-0132484P.

PR 06-MAY-1999; 99US-0132485P.

PR 06-MAY-1999; 99US-0132486P.

PR 07-MAY-1999; 99US-0132863P.

PR 11-MAY-1999; 99US-0134256P.

PR 14-MAY-1999; 99US-0134218P.

PR 14-MAY-1999; 99US-0134219P.

PR 14-MAY-1999; 99US-0134221P.

PR 18-MAY-1999; 99US-0134370P.

PR 19-MAY-1999; 99US-0134688P.

PR 20-MAY-1999; 99US-0134941P.

PR 21-MAY-1999; 99US-0135124P.

PR 24-MAY-1999; 99US-0135533P.

PR 25-MAY-1999; 99US-0135629P.

PR 27-MAY-1999; 99US-0136021P.

PR 28-MAY-1999; 99US-0136392P.

PR 01-JUN-1999; 99US-0136782P.

PR 03-JUN-1999; 99US-0137222P.

PR 04-JUN-1999; 99US-0137528P.

PR 07-JUN-1999; 99US-0137724P.

PR 08-JUN-1999; 99US-0138094P.

PR 10-JUN-1999; 99US-0138540P.

PR 10-JUN-1999; 99US-0138847P.

PR 14-JUN-1999; 99US-0139119P.

PR 16-JUN-1999; 99US-0139452P.

PR 16-JUN-1999; 99US-0139453P.

PR 17-JUN-1999; 99US-0139454P.

PR 18-JUN-1999; 99US-0139454P.

PR 18-JUN-1999; 99US-0139455P.

PR 18-JUN-1999; 99US-0139456P.

PR 18-JUN-1999; 99US-0139457P.

PR 18-JUN-1999; 99US-0139458P.

PR 18-JUN-1999; 99US-0139459P.

PR 18-JUN-1999; 99US-0139460P.

PR 18-JUN-1999; 99US-0139461P.

PR 18-JUN-1999; 99US-0139462P.

PR 18-JUN-1999; 99US-0139463P.

PR 18-JUN-1999; 99US-0139750P.

PR 18-JUN-1999; 99US-0139763P.

PR 21-JUN-1999; 99US-0139817P.

PR 22-JUN-1999; 99US-0139899P.

PR 23-JUN-1999; 99US-0140353P.

PR 23-JUN-1999; 99US-0140354P.

PR 24-JUN-1999; 99US-0140635P.

PR 28-JUN-1999; 99US-0140823P.

PR 29-JUN-1999; 99US-0140991P.

PR 30-JUN-1999; 99US-0141287P.

PR 01-JUL-1999; 99US-0141842P.

PR 01-JUL-1999; 99US-0142154P.

PR 02-JUL-1999; 99US-0142055P.

PR 06-JUL-1999; 99US-0142390P.

PR 08-JUL-1999; 99US-0142803P.

PR 09-JUL-1999; 99US-0142920P.

PR 12-JUL-1999; 99US-0142977P.

PR 13-JUL-1999; 99US-0143542P.

PR 14-JUL-1999; 99US-0143624P.

PR 15-JUL-1999; 99US-0144005P.

PR 16-JUL-1999; 99US-0144085P.

PR 16-JUL-1999; 99US-0144086P.

PR 19-JUL-1999; 99US-0144325P.

PR 19-JUL-1999; 99US-0144331P.

PR 19-JUL-1999; 99US-0144332P.

PR 19-JUL-1999; 99US-0144333P.

PR 19-JUL-1999; 99US-0144334P.

PR 19-JUL-1999; 99US-0144335P.

PR 20-JUL-1999; 99US-0144352P.

PR 20-JUL-1999; 99US-0144632P.

PR 20-JUL-1999; 99US-0144884P.

PR 21-JUL-1999; 99US-0144814P.

PR 21-JUL-1999; 99US-0145086P.

PR 22-JUL-1999; 99US-0145087P.

PR 22-JUL-1999; 99US-0145087P.

PR 22-JUL-1999; 99US-0145089P.

PR 22-JUL-1999; 99US-0145192P.

PR 23-JUL-1999; 99US-0145145P.

PR 23-JUL-1999; 99US-0145218P.

PR 23-JUL-1999; 99US-0145222P.

PR 26-JUL-1999; 99US-0145276P.

PR 27-JUL-1999; 99US-0145913P.

PR 27-JUL-1999; 99US-0145918P.

PR 27-JUL-1999; 99US-0145919P.

PR 28-JUL-1999; 99US-0145951P.

PR 02-AUG-1999; 99US-0146386P.

PR 02-AUG-1999; 99US-0146388P.

PR 02-AUG-1999; 99US-0146389P.

PR 03-AUG-1999; 99US-0147038P.

PR 04-AUG-1999; 99US-0147204P.

PR 05-AUG-1999; 99US-0147302P.

PR 05-AUG-1999; 99US-0147192P.

PR 05-AUG-1999; 99US-0147260P.

PR 06-AUG-1999; 99US-0147303P.

PR 06-AUG-1999; 99US-0147416P.

PR 09-AUG-1999; 99US-0147493P.

PR 09-AUG-1999; 99US-0147935P.

PR 10-AUG-1999; 99US-0148171P.

PR 11-AUG-1999; 99US-0148319P.

PR 12-AUG-1999; 99US-0148341P.

PR 13-AUG-1999; 99US-0148565P.

PR 13-AUG-1999; 99US-0148684P.

PR 16-AUG-1999; 99US-0149368P.

PR 17-AUG-1999; 99US-0149175P.

PR 18-AUG-1999; 99US-0149426P.

PR 20-AUG-1999; 99US-0149722P.

PR 20-AUG-1999; 99US-0149723P.

PR 20-AUG-1999; 99US-0149929P.

PR 23-AUG-1999; 99US-0149902P.

PR 23-AUG-1999; 99US-0149930P.

PR 25-AUG-1999; 99US-0150566P.

PR 26-AUG-1999; 99US-0150884P.

PR 27-AUG-1999; 99US-0151065P.

PR 27-AUG-1999; 99US-0151066P.

PR 27-AUG-1999; 99US-0151080P.

PR 30-AUG-1999; 99US-0151303P.

PR 31-AUG-1999; 99US-0151338P.


```
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158332P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 26-OCT-1999; 99US-0161362P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.
```

Query Match 6.6%; Score 33.2; DB 3; Length 1417;

Best Local Similarity 61.6%; Pred. No. 9.2;

Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

```
QY 337 CTCACATAAAGTGTCCATCTGAGATGTCACAGAGGCTTGGGCAAGACAGAG 396  
DB 358 CTCACAACTTCAGTACTTGTGAGAAACAAAGGCGTTTCCCAATAGCAGAG 299  
QY 397 GGAAGCACTGTTGTGTGTTAAAG 422  
DB 298 AGGCGTCAACGTTGGCGTTGTAGAG 273
```

RESULT 15

AAC46837/c
ID AAC46837 standard; DNA; 1431 BP.

XX AAC46837;

DT 18-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 51586.

DE Hybridisation assay; generic mapping; gene expression control;

KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.

```
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 200SEP-00301439.  
XX  
PR 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139482P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.
```

PR	28-JUN-1999;	99US-0140991P
PR	30-JUN-1999;	99US-0141287P
PR	01-JUL-1999;	99US-0141842P
PR	01-JUL-1999;	99US-0142154P
PR	02-JUL-1999;	99US-0142055P
PR	02-JUL-1999;	99US-0142309P
PR	06-JUL-1999;	99US-0142803P
PR	08-JUL-1999;	99US-0142920P
PR	09-JUL-1999;	99US-0142977P
PR	12-JUL-1999;	99US-0143542P
PR	13-JUL-1999;	99US-0143624P
PR	14-JUL-1999;	99US-0144005P
PR	15-JUL-1999;	99US-0144085P
PR	16-JUL-1999;	99US-0144088P
PR	16-JUL-1999;	99US-0144332P
PR	19-JUL-1999;	99US-0144333P
PR	19-JUL-1999;	99US-0144333P
PR	19-JUL-1999;	99US-0144333P
PR	21-JUL-1999;	99US-0145086P
PR	21-JUL-1999;	99US-0145088P
PR	22-JUL-1999;	99US-0145085P
PR	22-JUL-1999;	99US-0145087P
PR	22-JUL-1999;	99US-0145089P
PR	22-JUL-1999;	99US-0145192P
PR	23-JUL-1999;	99US-0145145P
PR	23-JUL-1999;	99US-0145218P
PR	23-JUL-1999;	99US-0145248P
PR	23-JUL-1999;	99US-0145248P
PR	26-JUL-1999;	99US-0145376P
PR	26-JUL-1999;	99US-0145513P
PR	27-JUL-1999;	99US-0145518P
PR	27-JUL-1999;	99US-0145518P
PR	28-JUL-1999;	99US-0145511P
PR	02-AUG-1999;	99US-0146186P
PR	02-AUG-1999;	99US-0146388P
PR	03-AUG-1999;	99US-0146589P
PR	03-AUG-1999;	99US-0147038P
PR	04-AUG-1999;	99US-0147040P
PR	04-AUG-1999;	99US-0147302P
PR	05-AUG-1999;	99US-0147192P
PR	05-AUG-1999;	99US-0147260P
PR	06-AUG-1999;	99US-0147303P
PR	06-AUG-1999;	99US-0147416P
PR	09-AUG-1999;	99US-0147933P
PR	09-AUG-1999;	99US-0147935P
PR	10-AUG-1999;	99US-0148171P
PR	11-AUG-1999;	99US-0148319P
PR	12-AUG-1999;	99US-0148341P
PR	13-AUG-1999;	99US-0148365P
PR	13-AUG-1999;	99US-0148684P
PR	17-AUG-1999;	99US-0149368P
PR	17-AUG-1999;	99US-0149375P
PR	18-AUG-1999;	99US-0149426P
PR	18-AUG-1999;	99US-0149724P
PR	20-AUG-1999;	99US-0149723P
PR	20-AUG-1999;	99US-0149929P
PR	23-AUG-1999;	99US-0149902P
PR	23-AUG-1999;	99US-0149930P
PR	27-AUG-1999;	99US-0151068P
PR	27-AUG-1999;	99US-0151068P
PR	27-AUG-1999;	99US-0151303P
PR	31-AUG-1999;	99US-0151438P
PR	01-SEP-1999;	99US-0151930P
PR	07-SEP-1999;	99US-0153073P
PR	10-SEP-1999;	99US-0153073P

PR	13-SEP-1999;	99US-0153758P.
PR	15-SEP-1999;	99US-0154018P.
PR	16-SEP-1999;	99US-0154039P.
PR	20-SEP-1999;	99US-0154779P.
PR	22-SEP-1999;	99US-0155139P.
PR	23-SEP-1999;	99US-0155486P.
PR	24-SEP-1999;	99US-0155659P.
PR	28-SEP-1999;	99US-0156458P.
PR	29-SEP-1999;	99US-0156596P.
PR	04-OCT-1999;	99US-0157117P.
PR	05-OCT-1999;	99US-0157753P.
PR	07-OCT-1999;	99US-0157865P.
PR	08-OCT-1999;	99US-0158029P.
PR	08-OCT-1999;	99US-0158232P.
PR	12-OCT-1999;	99US-0158369P.
PR	13-OCT-1999;	99US-0159233P.
PR	13-OCT-1999;	99US-0159294P.
PR	13-OCT-1999;	99US-0159235P.
PR	14-OCT-1999;	99US-0159329P.
PR	14-OCT-1999;	99US-0159330P.
PR	14-OCT-1999;	99US-0159331P.
PR	14-OCT-1999;	99US-0159637P.
PR	14-OCT-1999;	99US-0159638P.
PR	18-OCT-1999;	99US-0159584P.
PR	21-OCT-1999;	99US-0160741P.
PR	21-OCT-1999;	99US-0160767P.
PR	21-OCT-1999;	99US-0160768P.
PR	21-OCT-1999;	99US-0160770P.
PR	21-OCT-1999;	99US-0160814P.
PR	21-OCT-1999;	99US-0160815P.
PR	22-OCT-1999;	99US-0160980P.
PR	22-OCT-1999;	99US-0160981P.
PR	22-OCT-1999;	99US-0160989P.
PR	25-OCT-1999;	99US-0161404P.
PR	25-OCT-1999;	99US-0161405P.
PR	25-OCT-1999;	99US-0161406P.
PR	26-OCT-1999;	99US-0161359P.
PR	26-OCT-1999;	99US-0161360P.
PR	28-OCT-1999;	99US-0161361P.
PR	28-OCT-1999;	99US-0161920P.
PR	28-OCT-1999;	99US-0161922P.
PR	29-OCT-1999;	99US-0161933P.
PR	29-OCT-1999;	99US-0162142P.

```

Query Match          6.6%; Score 33.2; DB 3; Length 1431;
Best Local Similarity 61.6%; Pred. No. 9.2;
Matches 53; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY      337  CTCACATAACGATGATGCCATCTCGAAGATGCACAGAGCCTTTGGGCGCAGAAGCAGAG 396
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db       373  CTTCAACAACCTTCAGTGACTTTGTGGAGCAACAAAGGCGTTTCCCAATAGCAGAG 314
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY      397  GGAAGGCACTGTTTGTGTGTTAAG 422
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db       313  AGGCGTCACGTTGGCGTTGTAGAG 288
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Search completed: February 10, 2005, 09:41:10
Job time : 348.632 secs

```

Search completed: February 10, 2005, 09:41:10

Job time : 348.632 secs

PR	0/-SEP-1999;	99US-0134303F.
PR	10-SEP-1999;	99US-0153070P.

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 22:08:58, Search time 103.92 Seconds
(without alignments)
7888.487 Million cell updates/sec

Title: US-09-926-375B-7_COPY_11500_12000

Perfect score: 501
Sequence: 1 aaccacacagacatcttc.....gactcatcgtggtctgttt 501

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Issued Patents NA:
1: /cgn2_6/pcodaca/1/ina/5A_COMB.seq.*
2: /cgn2_6/pcodaca/1/ina/5B_COMB.seq.*
3: /cgn2_6/pcodaca/1/ina/6A_COMB.seq.*
4: /cgn2_6/pcodaca/1/ina/6B_COMB.seq.*
5: /cgn2_6/pcodaca/1/ina/PCUS_COMB.seq.*
6: /cgn2_6/pcodaca/1/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	35.6	7.1	152466	4 US-09-949-016-12869	Sequence 12869, A
C 2	34.4	6.9	390890	4 US-09-949-016-14720	Sequence 14720, A
C 3	34.2	6.8	144362	4 US-09-949-016-16066	Sequence 16066, A
C 4	33.8	6.7	1055	4 US-09-806-7088-23	Sequence 23, Appl
C 5	33.4	6.7	101674	4 US-09-949-016-12033	Sequence 12033, A
C 6	33.3	6.6	27250	4 US-09-949-016-15537	Sequence 15537, A
C 7	32.6	6.5	601	4 US-09-949-016-64253	Sequence 64253, A
C 8	32.6	6.5	183202	4 US-09-949-016-13614	Sequence 13614, A
C 9	32.4	6.5	832	4 US-09-621-976-2813	Sequence 2813, Ap
C 10	32	6.4	91665	4 US-09-949-016-12234	Sequence 12234, A
C 11	32	6.4	134890	4 US-09-949-016-15602	Sequence 15602, A
C 12	31.4	6.3	786431	4 US-09-751-389-3	Sequence 3, Appl
C 13	31.2	6.2	222452	4 US-09-949-016-12968	Sequence 12968, A
C 14	31	6.2	2109	4 US-09-949-016-5757	Sequence 5757, Ap
C 15	31	6.2	26011	4 US-09-949-016-17499	Sequence 17499, A
C 16	31	6.2	94133	4 US-09-949-016-11901	Sequence 11901, A
C 17	31	6.2	94133	4 US-09-949-016-12713	Sequence 12713, A
C 18	31	6.2	94135	4 US-09-949-016-15934	Sequence 15934, A
C 19	31	6.2	94135	4 US-09-949-016-15935	Sequence 15935, A
C 20	31	6.2	94135	4 US-09-949-016-15936	Sequence 15936, A
C 21	31	6.2	94135	4 US-09-949-016-15937	Sequence 15937, A
C 22	30.8	6.1	601	4 US-09-949-016-17807	Sequence 17807, A
C 23	30.8	6.1	98302	4 US-09-328-475C-19	Sequence 19, Appl
C 24	30.6	6.1	980	4 US-09-949-016-15250	Sequence 15250, A
C 25	30.6	6.1	21679	4 US-09-949-016-13356	Sequence 13356, A
C 26	30.6	6.1	31820	4 US-09-949-016-13356	Sequence 13356, A
C 27	30.6	6.1	232024	4 US-09-949-016-13477	Sequence 13477, A

C 28	30.6	6.1	784019	4 US-09-949-016-14033	Sequence 14033, A
C 29	30.6	6.1	828152	4 US-09-949-016-12777	Sequence 12777, A
C 30	30.4	6.1	402	4 US-09-583-110-842	Sequence 842, App
C 31	30.4	6.1	38368	4 US-09-949-016-12958	Sequence 12958, A
C 32	30.2	6.0	256	1 US-08-616-368A-25	Sequence 25, Appl
C 33	30.2	6.0	256	3 US-09-054-298-25	Sequence 25, Appl
C 34	30.2	6.0	256	3 US-08-818-655-25	Sequence 25, Appl
C 35	30.2	6.0	256	4 US-09-305-839-25	Sequence 25, Appl
C 36	30.2	6.0	601	4 US-09-949-016-55292	Sequence 55292, A
C 37	30.2	6.0	107329	4 US-09-949-016-12663	Sequence 12663, A
C 38	30.2	6.0	107329	4 US-09-949-016-12664	Sequence 12664, A
C 39	30.2	6.0	107330	4 US-09-949-016-15408	Sequence 15408, A
C 40	30.2	6.0	107330	4 US-09-949-016-15409	Sequence 15409, A
C 41	30.2	6.0	107330	4 US-09-949-016-15410	Sequence 15410, A
C 42	30.2	6.0	107330	4 US-09-949-016-15411	Sequence 15411, A
C 43	30.2	6.0	107750	4 US-09-949-016-12662	Sequence 12662, A
C 44	30.2	6.0	107751	4 US-09-949-016-15412	Sequence 15412, A
C 45	30.2	6.0	107751	4 US-09-949-016-15413	Sequence 15413, A

ALIGNMENTS

```

RESULT 1
US-09-949-016-12869/C
; Sequence 12869, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTNER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241, 755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237, 768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231, 498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12869
; LENGTH: 152486
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12869
Query Match      7.1% Score 35.6; DB 4; Length 152486;
Best Local Similarity 62.2%; Pred. No. 3;
Matches 56; Conservative 0; Mismatches 34; Indels 0; Gaps 0;
QY      23 TTGATTCATTCGATCTTCTTTAGACATGACACTGACCCCTGTGGCACTGC 82
Db      113516 TTTTCATCTTCGTATATTCCTTCACATCCTTCACATCGGCTGTTCTTCTT 113457
QY      83 AAGTTGACACATPAGCGCTTTGAACCTGG 112
Db      113456 CTGTTGCTAAAGGGGCACTTCCTTGG 113427

RESULT 2
US-09-949-016-14720/C
; Sequence 14720, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTNER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241, 755

```

```

; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14720
; LENGTH: 390890
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(390890)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14720

```

```

Query Match
Best Local Similarity 6.9%; Score 34.4; DB 4; Length 390890;
Matches 74; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

```

```

QY 12 ACATCTTCATGATTCATTCGATCTTCTTTAGACAATAGTGAACCCCTTG 71
DB 364776 AGATTTTATTTCCAGACTCTTAAGACTTCTCTCAGATCAGATGCACTGATTCAT 364717
QY 72 TTGGCAGTCTCAAGTTTGAACAACATAGGCGCTTTGAACCTTGGCAAGGTCACTGTC 131
DB 364716 TTGGTATCTGATCTTAAGAAAGGTGGGCTTAAACCTAGATGAGGATGATGATGTC 364657
QY 132 ACCCAAGCATCTCGGTGAC 151
DB 364656 AGCAACACACCATGAGCAC 364637

```

```

RESULT 3
US-09-949-016-16066/c
; Sequence 16066, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16066
; LENGTH: 144362
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(144362)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16066

```

```

Query Match
Best Local Similarity 6.8%; Score 34.2; DB 4; Length 144362;
Matches 75; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

```

```

QY 2 ACCCAAGCAATCTTTCATTCATTCGATCTTCTTTAGACAATAGTGAAC 61
DB 27696 ACACCCCTTACTTACTTACCGTAATCTGTTGGCTCTACATGAAACATACCAAT 27637
QY 62 TGACCCCTTGTGGCAGTCTCAAGTTTGACAACATAGGCGCTTTGAACTTGGCAAGGTC 121

```

```

DB 27636 TCAACCACTTCTCTCAAACTCACTGTATTACCACTTGTGTAGCACTTCATTCCTC 27577
QY 122 CATCACTGTACCCAGCATCT 144
DB 27576 CTAAATTAACAACAAGCTCTCT 27554

```

```

RESULT 4
US-09-806-708B-23
; Sequence 23, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; PRIOR FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 23
; LENGTH: 1055
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)...(1055)
; OTHER INFORMATION: consensus sequence of A.t. and L.a. PAB1 promoters
US-09-806-708B-23

```

```

Query Match
Best Local Similarity 6.7%; Score 33.8; DB 4; Length 1055;
Matches 47; Conservative 47; Mismatches 87; Indels 0; Gaps 0;

```

```

QY 1 AACCCAGCAATCTTCCATTCATTCGATCTTCTTTAGACAATAGTGAAC 60
DB 303 AAMVAAAMMMWSATTWMAAAMTSRTWTYGHKTMNNNNNGTTCWRMAWTKMMK 362
QY 61 CTGACCCCTTGTGGCAGTCTCAAGTTTGACAACATAGGCGCTTTGAACCTTGGCAAGGT 120
DB 363 TGGTTNNNGRTYGTWTKMATTYAKNNCTTAAMKKTCTMNTTAAKATTYAT 422
QY 121 CCATCAGTGCACCCAGCATCTCGGTGACCTTGGTGAATATCTGGCTAACCTT 180
DB 423 CYKXMTGTSYRYAALATYAWWTRRYAANNNTTWKACGWTYKSCCTANNTA 482
QY 181 A 181
DB 483 A 483

```

```

RESULT 5
US-09-949-016-12033
; Sequence 12033, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12033
; LENGTH: 101674

```

TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)... (101674)
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12033

Query Match
Best Local Similarity 58.6%; Pred. No. 13;
Matches 58; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 307 TGCATCTTCTTCAAGCAGTGTGTAATCTCACTAATGATGCTTCCATCTGAGCT 366
DB 77931 TTCTTTTATTATACAGCAGGCACTTAATCCACACTGATGAGTTCTTTTGACAG 77990
QY 367 CCACGAGGCTTTGGGCAAGAGAGAGGAGGAC 405
DB 77991 CCATTAAAGAGAGGAGGAGTGAAGAGGTTGAAAACC 78029

RESULT 6
US-09-949-016-15537
Sequence 15537, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 15537
LENGTH: 27250
TYPE: DNA
ORGANISM: Human
US-09-949-016-15537

Query Match
Best Local Similarity 53.5%; Pred. No. 9;
Matches 69; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 125 CACTGTCAACCAAGCATCTGGTGACCTTGGTTGGAATATCTGGCTAATCTAGAT 184
DB 25818 CACAGTCTCCCAAGTACTGACTACAGGTGTCTACCAACATCCCTGTAATTTTAG 25877
QY 185 ATTTTCTTGGAGTATCTTTAGAACATCCAGAAATAGGCTGATTTCTATCTCTGGAC 244
DB 25878 TTTTGTGTTTATATATTATTTATTTTAAAGAACAGATCTCCTGCAACCCAGGC 25937
QY 245 CACATATTA 253
DB 25938 TAGAGTACA 25946

RESULT 7
US-09-949-016-64253/C
Sequence 64253, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016

CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 64253
LENGTH: 601
TYPE: DNA
ORGANISM: Human
US-09-949-016-64253

Query Match
Best Local Similarity 49.7%; Pred. No. 1,6;
Matches 83; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 228 GATTCATCTCTGGGACCAATATTAATCCCTAGATCCAGAGATCGTCAGAGA 287
DB 227 GACCATCAAAATGAGAAAAAACAATTAATACAGACTTACGTGATTGTTCAAGAG 168
QY 288 AACAGATCTCTCTGCTGTCATCTTCTCAAGCAGTGAATGATCTCCATTAAC 347
DB 167 ATTAATATGCAAAATGCTTGTTTAAAAATGCAAGGAAAGTAACTGTACTTAA 108
QY 348 TGAGTCCCATCTGAGATCCACAGAGGCTTTGGGCAAGAACG 394
DB 107 TGAGACACTTTATATAGAGAGAGAAAGATTGGCGTAAAGATGAG 61

RESULT 8
US-09-949-016-13614
Sequence 13614, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13614
LENGTH: 183202
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)... (183202)
OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13614

Query Match
Best Local Similarity 49.7%; Pred. No. 33;
Matches 83; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 228 GATTCATCTCTGGGACCAATATTAATCCCTAGATCCAGAGATCGTCAGAGA 287
DB 11606 GACCATCAAAATGAGAAAAAACAATTAATACAGACTTACGTGATTGTTCAAGAG 116065
QY 288 AACAGATCTCTCTGCTGTCATCTTCTCAAGCAGTGAATGATCTCCATTAAC 347
DB 11606 ATTAATATGCAAAATGCTTGTTTAAAAATGCAAGGAAAGTAACTGTACTTAA 116125

Qy 348 TGAGTTCCTCATCTGAGAGTCACAGAGGCTTTGGGGCAAGAGCAG 394
116126 TGAGGACACTTTTATGAGAGAGAGAAAGATTGGCGGTAAAGATGAG 116172

RESULT 9
US-09-621-976-2813
Sequence 2813, Application US/09621976
Patent No. 6639063

GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Jobert, S.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: ESTs and Encoded Human Proteins.
FILE REFERENCE: GENSET.054PR2
CURRENT APPLICATION NUMBER: US/09/621,976

CURRENT FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 19335
SOFTWARE: Patent.pm
SEQ ID NO 2813
LENGTH: 832
TYPE: DNA

ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 235..399
US-09-621-976-2813

Query Match 6.5%; Score 32.4; DB 4; Length 832;
Best Local Similarity 9.9%; Pred. No. 2.2; Indels 1; Gaps 1;
Matches 35; Conservative 166; Mismatches 152;

Qy 86 TTGACAACTATAGGCTTTGAACTTGACCAAGTCCATCATGCCCAAGCATCTG 145
10 TWAKCMTKMKMSYMWYMKYMKYMKYMKYMKYMKYMKYMKYMKYMKYMKYMKY 69
Qy 146 GGTGACCTTTGGTGAATATTTGGCTACTT-TAGATATTTCTTGGAGTATCTT 204
70 KKAMCRTKTKKK 129
Qy 205 AGAACAATCCAGAAATAGGCTTGATCTCATCTGAGCAACAAATATAAGTCCCTG 264
130 YRKTTCYSSKGTWYMKYMKYMKYMKYMKYMKYMKYMKYMKYMKYMKYMKY 189
Db 130 YRKTTCYSSKGTWYMKYMKYMKYMKYMKYMKYMKYMKYMKYMKYMKYMKY 189
Qy 265 AATCCAGAGATCGTGACAGAGAAAGAGATCTCTCTGTCATCTCTTCTCAAGC 324
190 ARKMTSTWRKRSYASASARCCYSCWGMAMSKYMKYMKYMKYMKYMKYMKY 249
Qy 325 AGTGAGTAGTACTCCCACTAACTGAGTCCCATCTGAGAGTCCACAGAGGCTTGGGG 384
250 RRRYAGSKSYKSYKSYKSYKSYKSYKSYKSYKSYKSYKSYKSYKSYKSYK 309
Qy 385 CAAGAACAGAGGAGGACCTGTTGTGTGTAAGTTTGAAGCTTAACTTCAAA 438
310 WMMCMAMMYSTCTRASMWRMYIYMMKMKYMAAABAAMWMAWMAWMAWMA 363
Db 310 WMMCMAMMYSTCTRASMWRMYIYMMKMKYMAAABAAMWMAWMAWMAWMA 363

RESULT 10
US-09-949-016-12234/C
Sequence 12234, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: CLO01307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498

PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12234
LENGTH: 91665
TYPE: DNA
ORGANISM: Human
US-09-949-016-12234

Query Match 6.4%; Score 32; DB 4; Length 91665;
Best Local Similarity 45.3%; Pred. No. 37;
Matches 116; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

Qy 60 ACTGACCCCTTTGGGAGCTCTCAAGTTTGACAACTATAGGCTTTGAAGTGGACCAAG 119
9823 AGTAATCAGTCTGACTAGGCCCAAGGATGTGTATATTTTGCATTTGAAATATAT 9764
Qy 120 TCCATCACTGTACCCCAAGCATCTCGGTGACCTTTGGTTGAATATCTTGGCTAACT 179
Db 9763 ATCTTAATGAGCAAGAAATTTAGTAGATATCTTGTAGCAGCTCTGGTTAACT 9704
Qy 180 TAGATATTTCTTTGGAGTATCTTTAGAACTCCAGAAATAGGCTTTGATCTCATCCT 239
Db 9703 CAGAGGAATTAAGGCGCTGTCTAAATATGAGTTGTATATGACAGGTAAGAAAT 9644
Qy 240 GGCACCAATATATAGTCACTGAAATCCAGAGATCGTGACAGAAACAAGATCTC 299
Db 9643 GCACCCCAAAATATGCACTTTGGAATTTGAAATATGAGAAATAGCAAGAAAGGTTTCTC 9584
Qy 300 TCTGCTGTCATCTT 315
Db 9583 TGACCTCTTGCTCCTT 9568

RESULT 11
US-09-949-016-15602
Sequence 15602, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: CLO01307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 15602
LENGTH: 134890
TYPE: DNA
ORGANISM: Human
US-09-949-016-15602

Query Match 6.4%; Score 32; DB 4; Length 134890;
Best Local Similarity 60.2%; Pred. No. 45;
Matches 53; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

Qy 84 AGTTGACAACTATAGGCTTTGAAGTGGACAGAGTCCATCATGCCCAAGCATCC 143
Db 89981 ATTTATGAAAGAGATTGGAGCTGGGCAATGGCTCATCTGTATATCCAGACTT 90040
Qy 144 TGGGTGACCTTTGGGTTGGAATATCTTG 171
Db 90041 TGGGAGCCCAAGGAAGATGATCATCTTG 90068

RESULT 12

ORGANISM: Human
US-09-949-016-17499

Query Match 6.2%; Score 31; DB 4; Length 26011;
Best Local Similarity 59.8%; Pred. No. 41;

Matches 52; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

Qy	260	CCTAGATCCCAAGAGATCGTCAGAGAAACAAGATCTCTCGTGCATCCTTTC	319
Db	23278	CATAGACGACATGAAGAACAGCAGTGAACCAAGATTCTAGCAAGGCCAGCTATTAGC	23219
Qy	320	AAAGCAGTGAGTAGTCTCACTAA	346
Db	23218	AAAGCAGTAGAGGAACTGAACTAGA	23192

Search completed: February 10, 2005, 17:42:54
Job time : 108.92 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 9, 2005, 22:13:58 ; Search time 365.063 Seconds
(without alignments)
8097.594 Million cell updates/sec

Title: US-09-926-375B-7_COPY_11500_12000

Perfect score: 501
Sequence: 1 aaccacacgaacatctctc.....gactcatgtgggtctctt 501

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 5378673 seqs, 295022984 residues

Total number of hits satisfying chosen parameters: 10757346

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database:

Published Applications NA:
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
20: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
21: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
22: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	37	7.4	3555	US-10-282-122A-16204	Sequence 16204, A
2	36.8	7.3	49088	US-10-087-192-13	Sequence 13, Appl
3	36.8	7.3	56773	US-10-331-053-47	Sequence 47, Appl
4	36	7.2	503	US-10-027-632-138049	Sequence 138049, A
5	36	7.2	503	US-10-027-632-138049	Sequence 138049, A
6	36	7.2	1032	US-10-424-599-132734	Sequence 132734, A
7	35.4	7.1	34511	US-10-085-117-307	Sequence 307, Appl
8	34.8	6.9	51806	US-10-719-993-6892	Sequence 6892, App
9	34.6	6.9	4352	US-10-741-600-776	Sequence 776, App
10	34.4	6.9	634	US-10-424-599-80920	Sequence 80920, A
11	34.4	6.9	4881	US-10-437-963-40327	Sequence 40327, A

C 12	34.4	6.9	518360	18	US-10-367-094-125	Sequence 125, App
C 13	34.2	6.8	2859	13	US-10-027-632-264212	Sequence 264212, A
C 14	34.2	6.8	2859	17	US-10-027-632-264212	Sequence 264212, A
C 15	33.6	6.7	295096	13	US-10-087-192-331	Sequence 331, App
C 16	33.4	6.7	1536	17	US-10-424-599-120851	Sequence 120851, A
C 17	32.8	6.5	599	13	US-10-027-632-280124	Sequence 280124, A
C 18	32.8	6.5	599	17	US-10-027-632-280124	Sequence 280124, A
C 19	32.6	6.5	311	11	US-09-864-408A-6439	Sequence 6439, App
C 20	32.6	6.5	321	18	US-10-674-124A-2052	Sequence 2052, A
C 21	32.6	6.5	508	13	US-10-027-632-141634	Sequence 141634, A
C 22	32.6	6.5	508	17	US-10-027-632-141634	Sequence 141634, A
C 23	32.6	6.5	89829	18	US-10-322-281-359	Sequence 359, App
C 24	32.6	6.5	197526	18	US-10-322-281-359	Sequence 359, App
C 25	32.4	6.5	1462	17	US-10-424-599-87448	Sequence 87448, A
C 26	32.4	6.5	1793	17	US-10-425-114-29850	Sequence 29850, A
C 27	32.4	6.5	1816	17	US-10-425-114-29850	Sequence 29850, A
C 28	32.2	6.4	315	9	US-09-783-590-1719	Sequence 1719, App
C 29	32.2	6.4	37925	11	US-09-984-429-701	Sequence 701, App
C 30	32.2	6.4	211257	13	US-10-087-192-529	Sequence 529, App
C 31	32	6.4	1808	17	US-10-369-493-33341	Sequence 33341, A
C 32	31.8	6.3	434	13	US-10-027-632-38203	Sequence 38203, A
C 33	31.8	6.3	434	17	US-10-027-632-38203	Sequence 38203, A
C 34	31.8	6.3	542	13	US-10-027-632-63825	Sequence 63825, A
C 35	31.8	6.3	542	13	US-10-027-632-65209	Sequence 65209, A
C 36	31.8	6.3	542	13	US-10-027-632-65209	Sequence 65209, A
C 37	31.8	6.3	542	17	US-10-027-632-65209	Sequence 65209, A
C 38	31.8	6.3	542	17	US-10-027-632-65209	Sequence 65209, A
C 39	31.8	6.3	542	17	US-10-027-632-296771	Sequence 296771, A
C 40	31.8	6.3	825	17	US-10-027-632-296771	Sequence 296771, A
C 41	31.8	6.3	3094	13	US-10-282-122A-16747	Sequence 16747, A
C 42	31.8	6.3	3094	17	US-10-027-632-177175	Sequence 177175, A
C 43	31.8	6.3	5969	15	US-10-027-632-177175	Sequence 177175, A
C 44	31.8	6.3	5969	17	US-10-172-066-45	Sequence 45, Appl
C 45	31.8	6.3	5969	18	US-10-240-588C-135	Sequence 135, Appl
					US-10-311-507-83	Sequence 83, Appl

ALIGNMENTS

RESULT 1
US-10-282-122A-16204
; Sequence 16204, Application US/10282122A
; Publication No. US20040029129A1
GENERAL INFORMATION:
; APPLICANT: Wang, Liangu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlson, Karl
; APPLICANT: Zyckind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: EPIITA.034A
CURRENT FILING DATE: 2003-02-20
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO: 16204
LENGTH: 3555
TYPE: DNA
ORGANISM: Clostridium botulinum
US-10-282-122A-16204

Query Match 7.4%; Score 37; DB 17; Length 3555;
Best Local Similarity 56.0%; Pred. No. 1.2; Indels 0; Gaps 0;
Matches 70; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

181 AGATATTTCTTTGAGATCTTTAGACATCCAGGAATAGGCTTGATTCATCTG 240
1947 AGTATTAATAGCAGATAGTTATATAGACAGCCCTATGGCTGAATACCAAGATA 2006
241 GGACCAATATATAGTACCCCTTGAATCCAGAGATCTGTGAGAAACAGATCTCT 300
2007 TATAGCTAATAAATTAACCTTTAATATAGATTCAAACGAAATCTGAACCTAT 2066
301 CTGCT 305
2067 ATTTGT 2071

RESULT 2

US-10-087-192-13/C
Sequence 13, Application US/10087192
Publication No. US20020182586A1
GENERAL INFORMATION:
APPLICANT: Morris, David W.
APPLICANT: Engelhard, Eric K.
TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
TITLE OF INVENTION: CANCER
FILE REFERENCE: 529452000122
CURRENT APPLICATION NUMBER: US/10/087,192
CURRENT FILING DATE: 2002-03-01
PRIOR APPLICATION NUMBER: US 09/747,377
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/798,586
PRIOR FILING DATE: 2001-03-02
NUMBER OF SEQ ID NOS: 2059
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13
LENGTH: 49088
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(49088)
OTHER INFORMATION: n = A,T,C or G
US-10-087-192-13

Query Match 7.3%; Score 36.8; DB 13; Length 49088;
Best Local Similarity 59.6%; Pred. No. 5.1; Indels 0; Gaps 0;
Matches 62; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

255 GTCACTAGATCCAGAGATGTCAGAGAAAGATCTCTCTGTCATCTCT 314
45517 GTCTCAAAACATCTGTGGAGTTGACCAAGTAAGATTAAGTTTAAACCCCTGAACT 45458
315 TCTTCAAGCAGTAGTAGTCTCACTCACTAACTGAAGTTCCCAT 358
45457 ACATTAAGCAGAAAGAAATGACCCCACTAAGTAGTCTTTCAT 45414

RESULT 3
US-10-331-053-47/C
Sequence 47, Application US/10331053
Publication No. US2004019778A1
GENERAL INFORMATION:
APPLICANT: David W. Morris
APPLICANT: Marc S. Malandro
TITLE OF INVENTION: Novel Compositions and Methods in Cancer
FILE REFERENCE: 529452001100
CURRENT APPLICATION NUMBER: US/10/331,053
CURRENT FILING DATE: 2002-12-26
NUMBER OF SEQ ID NOS: 86
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 47
LENGTH: 56773
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(56773)
OTHER INFORMATION: n = A,T,C or G
US-10-331-053-47

Query Match 7.3%; Score 36.8; DB 18; Length 56773;
Best Local Similarity 59.6%; Pred. No. 5.5; Indels 0; Gaps 0;
Matches 62; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

255 GTCACTAGATCCAGAGATGTCAGAGAAAGATCTCTCTGTCATCTCT 314
48995 GTCTCAAAACATCTGTGGAGTTGACCAAGTAAGATTAAGTTTAAACCCCTGAACT 48936
315 TCTTCAAGCAGTAGTAGTCTCACTCACTAACTGAAGTTCCCAT 358
48935 ACATTAAGCAGAAAGAAATGACCCCACTAAGTAGTCTTTCAT 48892

RESULT 4

US-10-027-632-138049
Sequence 138049, Application US/10027632
Publication No. US20020198371A1
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
TITLE OF INVENTION: Polymorphisms in the Human Genome
FILE REFERENCE: 109827,129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 138049
LENGTH: 503
TYPE: DNA
ORGANISM: Human
US-10-027-632-138049

Query Match 7.2%; Score 36; DB 13; Length 503;
Best Local Similarity 47.7%; Pred. No. 0.92; Indels 0; Gaps 0;
Matches 105; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

Qy	245	CACAAATATAGTACCCCTTGAATCCCAAGAGATCGTCAGAGAAA	CAAGATTCCTCTCG	304
Db	185	ACTGCCCTTTCCGGGGCTCTATGACCTGTAGAGAAACAGGGGCTGAGG	CACAGCCCATGAT	244
Qy	185	ATTTCCTTTGGAGATATCTTTTAAACAATCCAGAGAAA	TAGGGCTTGATCTCATCTCGGAC	244
Db	245	TTTCTGACACCTCTCCCTTAGAGTCTACCTCAAGGGTGCCTGCTTTCAT	ACCAAGCTCTCTCT	304
Qy	305	TGTGCATCTCTTTTCAAAGCAGTGAATGTA	CTCACTAAACTGAGTTCCCATCTGAGA	364
Db	305	GTTCCTCTCCCTTAAATGATGCCCTCTCTCGCCCTGACCAAGGAA	CTGTCTCTTTGTAC	364
Qy	365	GTCCACAGAGAGCTTTTGGGGCAAGAAAGCAGAGAGGAAAGCA	404	
Db	365	ATCTCTCAGGCTGAATGAGACAGAGAAAGCTCAAGGGGA	404	

```

RESULT 5
US-10-027-632-138049
: Sequence 138049, Application US/10027632
: Publication No.: US20030204075A9
: GENERAL INFORMATION:
: APPLICANT: Wang, David G.
: TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
: POLYMORPHISMS IN THE HUMAN GENOME
: FILE REFERENCE: 108827.129
: CURRENT APPLICATION NUMBER: US/10/027,632
: PRIOR FILING DATE: 2002-04-30
: PRIOR APPLICATION NUMBER: US 60/218, 006
: PRIOR FILING DATE: 2000-07-12
: PRIOR APPLICATION NUMBER: US 60/198, 676
: PRIOR FILING DATE: 2000-04-20
: PRIOR APPLICATION NUMBER: US 60/193, 483
: PRIOR FILING DATE: 2000-03-29
: PRIOR APPLICATION NUMBER: US 60/185, 218
: PRIOR FILING DATE: 2000-02-24
: PRIOR APPLICATION NUMBER: US 60/167, 363
: PRIOR FILING DATE: 1999-11-23
: PRIOR APPLICATION NUMBER: US 60/156, 358
: PRIOR FILING DATE: 1999-09-28
: PRIOR APPLICATION NUMBER: US 60/146, 002
: PRIOR FILING DATE: 1999-08-09
: NUMBER OF SEQ ID NOS: 325720
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 138049
: LENGTH: 503
: TYPE: DNA
: ORGANISM: Human
US-10-027-632-138049

```

Query Match	7.2%	Score 36;	DB 17;	Length 503;
Best Local Similarity	47.7%	Pred. No. 0.92;		
Matches 105; Conservative	0;	Mismatches 115;	Indels 0;	Gaps 0

QY 185 ATTTCTTTGGAAATATCTTTAGAAACATCCAGGAAATAGGGCTGTGATCTCATCCTGGAC 244

Db 185 ACTGCTTTGGCGGGGCTTAGACTCTAGAGGAACAGGGGCTGAGGCACAGCCCATGAT 244

QY 245 CACAAATATAGTACCCCTAGAAATCCAGAGAGATCTGTGACAGAAACAAGATCTCTCG 304

Db 245 TTTCTGACACCTCTCCTTAGAGTCTACTCTCAAGGGTGTGTCTTCATACAGCTCTCTCT 304

QY 305 TGTGCATCCTTTCTTTCAAGACGTGAGTAGTGATCTCCACTTAACTGAAGTTCCCATCTGAGA 364

Db 305 GTTTCCTTCCCTCTTAAGTGCCTCCCTCTCTGCTGACCAAGGAACTGCTCTTGTAC 364

QY 365 GTTCACAGAGGCTTTGGGCAAGAGACAGAGGGAAGGCA 404

Db 365 ATCTCAGGCTGAATGAGACACAGAGGAAGGTCAGAGGGA 404

RESULT 6

```

US-10-424-599-132734/C
; Sequence 132734, Application US/10424599
; Publication No. US20040031072A1
GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 132734
; LENGTH: 1032
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_90866C.1
US-10-424-599-132734

```

Query Match	7.2%;	Score 36;	DB 17;	Length 1032;
Best Local Similarity	55.6%;	Pred. No. 1.3;		
Matches 69;	Conservative 0;	Mismatches 55;	Indels 0;	Gaps 0;

[illegible]

```

RESULT 7
US-10-085-117-307/c
/ Sequence 307, Application US/10085117
/ Publication No. US2003023234A1
/
GENERAL INFORMATION:
/ APPLICANT: Morris, David W.
/ APPLICANT: Engelhard, Eric K.
/ TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
/ FILE REFERENCE: 529452000321
/
CURRENT APPLICATION NUMBER: US/10/085,117
CURRENT FILING DATE: 2002-02-27
PRIOR APPLICATION NUMBER: US 09/798,586
PRIOR FILING DATE: 2001-03-02
NUMBER OF SEQ ID NOS: 361
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 307
/
LENGTH: 34511
/
TYPE: DNA
/
ORGANISM: Mus musculus
/
FEATURE:
/ NAME/KEY: variation
/ LOCATION: (1)...(34511)
/
OTHER INFORMATION: n = any nucleotide
US-10-085-117-307

```

Query Match	7.1%	Score 35.4;	DB 17;	Length 34511;
Best Local Similarity	55.2%	Pred. No. 12;		
Matches	69;	Conservative	0;	Mismatches 56;
			Indels	0;
			Gaps	0

QY 212 GGAGATCTGTGAGGAAAACAAGATCTCTCTCGTGTCATCTTCTTCAAGAAGTAGT 311
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 31333 GGAGACTGTCCAGACAGTCAACCATCTCTGCATTTCATAAATTTTTCGAAGTAGTTC 3127

QY 332 AGTAGCTCACAATAACTGATGTTCCCATCTGAGATCCACAGAGGGCTTTGGGCAAGAA 391

Db 31273 TGTGCTTCTGATATGACGATTAATTTATCTTCTGAGTCTCTAGTGGGTTGAAG 31214
QY 392 CAGAG 396
Db 31213 ACTAG 31209

RESULT 8
US-10-719-993-6892/c
; Sequence 6892, Application US/10719993
; Publication No. US20040265849A1

; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6892
; LENGTH: 51806
; TYPE: DNA
; ORGANISM: Homo sapiens

US-10-719-993-6892

Query Match 6.9%; Score 34.8; DB 18; Length 51806;
Best Local Similarity 50.0%; Pred. No. 23;
Matches 87; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 316 CTTCAAGACGTGATGATGATCTCCACTAACTGAGTCCCATCTGAGAGTCACAGAG 375
Db 35334 CGTCAAGCGACAGTGTCTCTCAGTGAACAGATCAGAACTGAGGAAGTAGTGGCAGCTG 35275
QY 376 GCTTTGGGGCAGAGAGGAGGAGGACCTGTTTGTGTGAAGTTTGACTTAAC 435
Db 35274 GAATTGAGAAATACACAGACAAATGCTTAAGAGGACCCCTGATCTGTGAC 35215
QY 436 AAATTTGAGACATGATGATGATGCTGTCTGAGACTTAACAACCTTACTCATG 489
Db 35214 TAAATTTGAGCTGATCATATCATACAGAGGCTCTCAAAATTAACATAAG 35161

RESULT 9

US-10-741-600-776/c
; Sequence 776, Application US/10741600
; Publication No. US20050026169A1

; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001499
; CURRENT APPLICATION NUMBER: US/10/741,600
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 73997
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 776
; LENGTH: 4352
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(4352)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-741-600-776

Query Match 6.9%; Score 34.6; DB 19; Length 4352;
Best Local Similarity 53.3%; Pred. No. 7.8;
Matches 73; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 352 TTCCCATCTGAGATCCACAGAGGCTTTGGGCAAGAGCAGAGGAGGACTGTTTG 411

Db 4140 TTACATCATATTCACAAAGCAGGGTTACATGTAGAGAAAGAGAGGCTTCACTGC 4081
QY 412 TGTGTAAGTTTGACTCTTAACAATTGAAGACATAGATGATGTGTGAGACTAA 471
Db 4080 TTTTTCATCTGAGACTTAATGACCTGAATTAATTAATGAGAGTGTGTTTCTCA 4021
QY 472 CAACACCTAGACTCAT 488
Db 4020 AATGAATTACCAAT 4004

RESULT 10
US-10-424-599-80920

; Sequence 80920, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 80920
; LENGTH: 634
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_44087C.1
US-10-424-599-80920

Query Match 6.9%; Score 34.4; DB 17; Length 634;
Best Local Similarity 60.9%; Pred. No. 3.4;
Matches 56; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 142 CCTGGGTGACCTTGGTGGATATCTTGCTAACCTTAGATATTTCTTTGGAGTATC 201
Db 341 CCTCGGGAAATTTTGGGAGGAGATTTTGGTAAACCTTCATATGTTTTTTGAAACT 400
QY 202 TTAGAACAATCCAGAAATAGGGCTTGATCT 233
Db 401 ATTATGACATGCCGACCAAGTTGGGATTTT 432

RESULT 11

US-10-437-963-40327
; Sequence 40327, Application US/10437963
; Publication No. US2004012343A1

; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 40327
; LENGTH: 4881
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4510_43780C.1
US-10-437-963-40327

	Query Match	Similarity	6.9%	Score 38.4;	DB 18;	length 4881;
	Best Local	Similarity	56.0%;	Pred No. 9.6;	Mismatches	51; Indels 0; Gaps 0.
	Matches	Conservative	65;			
OY	324	CAGTGTGATGGACCTCCACTAACTGAATTCCCCTCATAGATGCCACAGAGCGTTTGSGG	383			
Dd	4088	CAGTTATTATGTGGCTATGCTATAACAAGSTAACCCTTAATGAGACTCTAAGTGGTATATG	4147			
OY	384	GCAAGAACAGAGGAGGAGGCACTGTTTTGTGTGTTAAATTGACTTCACCAAT	439			
Dd	4148	ATATGATCAAAGGGAATGCAGCACAATGTTGTTACATACCTTCATTAAATAAT	4203			

```

RESULT 12
US-10-367-094-125/c
; Sequence 125, Application US/10367094
; Publication No. US20040170982A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc Malandro
; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
; FILE REFERENCE: 529452001500
; CURRENT APPLICATION NUMBER: US/10/367, 094
; CURRENT FILING DATE: 2003-02-14
; NUMBER OF SEQ ID NOS: 203
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 125
; LENGTH: 518360
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1) .. (518360)
; OTHER INFORMATION: n = A,T,C or G
US-10-367-094-125

```

Query Match	6.9%	Score 34.4	DB 18	Length 518360
Best Local Similarity	52.9%	Pident. No. 1e+02	Mismatches 66	Indels 0
Matches 74	Conservative 0			Gaps 0
QY	12	ACATCTTCATGATTCATTCTGATCTCTTTCTAGACAATAGTGAACCCCTTG	71	
DB	480461	AGATTTTATTTCCAGACTCTAAGACTTCTCTCCAGATTCACAATAGCACTGATACCAT	4804020	
QY	72	TTGGCAGTTCGAAGTTTGACACATAGGGCTTTGAACTTGGCACAAGCTCATCACTGC	131	
DB	480401	TTCTGATATCTGATCTTAGAAGAGTGGGGCTTAAAACTAGATGACGAGTTGATAGGTGC	4803422	
QY	132	ACCCAGCATCTCGGTGAC	151	
DB	480341	AGCAAAACCACATGGCACAC	480322	

```

RESULT 13
; US-10-027-633-264212/C
; Sequence 264212, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotides
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24

```

```

? PRIOR APPLICATION NUMBER: US 60/167,363
? PRIOR FILING DATE: 1999-11-23
? PRIOR APPLICATION NUMBER: US 60/156,358
? PRIOR FILING DATE: 1999-09-28
? PRIOR APPLICATION NUMBER: US 60/146,002
? PRIOR FILING DATE: 1999-08-09
? NUMBER OF SEQ ID NOS: 32520
? SOFTWARE: FASTSEQ for Windows Version 4.0
? SEQ ID NO 264212
? LENGTH: 2859
? TYPE: DNA
? ORGANISM: Human
US-10-027-632-264212

```

US-10-027-632-264212

Query Match	6.8%;	Score 34.2;	DB 13;	Length 2659;
Best Local Similarity	50.3%;	Pred. No. 8.5;		
Matches	84;	Conservative	0;	Mismatches 83;
			Indels	0;
			Gaps	0;

Qy	155	GGGGTTGGAATACCTGGTACCTAGATATTTCTTGGAGATCCTTAGAACATCCA	214
Db	941	TGTGTGTGATTAATTAATCTAATCCCTCAATAGTTCCTCTGTGGAAATTAATACAGGTTTTTA	882
Qy	215	GGAAATAGGGCTTGATCTCATCTCTGGAGCCAAATATATAGTCACCTAGATCCAGGA	274
Db	881	TAACTTAAGCTTAATTTCATTTCTTAATTGACATTTTTCCTTACTTAAGACTTAGCA	822
Qy	275	GATCTGCAGAGAAACAAGATCTCTCTGTGTCATCCTTTCTCAA	321
Db	821	ACTCTTGAAGAAAAACAATGATGTGCTTAATTAGTAGCTATTAGAA	775

821 ACTCTTGAAAAACAATGATTGCTTATTAGTAGCTTATTAGAA 775

```

RESULT 14
US-10-027-632-264212/c
; Sequence 264212, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Mang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 264212
; LENGTH: 2859
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-264212

Query Match      6.8%; Score 34.2; DB 17; Length 2859;
Best Local Similarity 50.3%; Pred. No. 8.5;
Matches 84; Conservative 0; Mismatches 83; Indels 0; Gaps 0

155 TGGGTGGAATATCTGTGGTAACTTGAATATTTCTTTGGAGATCTTTGAAACATCA 214
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 941 TGGGTGGAATATTAATCTACTACCTCATCAATGTTCTCTGTGGAAATTAATACGTGTTTAA 882
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
215 GGAATATAGGCTTGATTTCTCATCTCTGGGACCAATATATAGCACCCATGAAATCCAGAA 274

```

DB 881 TAACATTAACCTTAATTTAATCTTAATGACATTTTCCCTTACTTAAGCTTAGCA 822
QY 275 GATCGTCAGAGAAACAAGATCTCTCGTGTGATCTTCTTCA 321
DB 821 ACTCTGAAAAAACAATGATGCTTATTTAGTAGCTTATTAGAA 775

RESULT 15
US-10-087-192-331/c
; Sequence 331, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhardt, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 52945200012
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 331
; LENGTH: 295096
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(295096)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-331

Query Match 6.7%; Score 33.6; DB 13; Length 295096;
Best Local Similarity 48.0%; Pred. No. 1.4e+02;
Matches 96; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 171 GGCTAACCTTAGATATTTCTTTGGAGATCTTTAGAACATCCAGAAATAGGGCTTGAT 230
DB 246768 GACTTTCTTGATGATCATTTAGACTAGATCTACCAACATTTGAAATGATGCCAGAT 246709
QY 231 TCTCATCTCGGACCAATATTAAGTCACTAGATCCCGAGATCGTGCAAGAAAC 290
DB 246708 TTTGAAAAATGGAACTAAAAAGGCCATTTTGAAGCAATTTAGCTCAAGTAAAAATA 246649
QY 231 AAGATCTCTCTCGTGTGATCTTCTTCAAGCAGTGAAGTAGTACTCCACTAACTGA 350
DB 246648 AAAAAATATGATGTCAGAAAACTCTCTGATTTGCTAAAGTGTCATCAACAGCTCTGA 246589
QY 351 GTTCCATCTGAGAGTCCAC 370
DB 246588 GTTGAAAACTCAACTCTC 246569

Search completed: February 10, 2005, 18:17:03
Job time : 371.063 secs